

What will happen to Russian science?

The Soviet Academy of Sciences has opted for allegiance to the Russian Republic rather than to what is left of the central government, but that by itself will not put things to rights.

It is not surprising that the Soviet Academy of Sciences, dragged by Stalin kicking and screaming from Leningrad (now again St Petersburg) to Moscow in the 1930s, should now have decided to throw in its lot with Mr Boris Yeltsin's Russian Republic (see page 690). Where else could it have gone? While what is left of the central government is still pleading with the republics for a treaty on economic collaboration that will enable it to pay its bills — eight of the fifteen, but not the Ukraine, did indeed sign up last Friday — nobody can be sure that there will ever again be the central funds to keep the academy in the style to which it has grown accustomed. But will Yeltsin's Russia be any safer a haven?

There are some obvious snags. One is that the academy may not be as thoroughly welcome as it would like to be. It has always been anomalous that the Russian republic differed from the others in not having an academy of its own. The all-union academy's view was that Russia did not need one, because the — 'big' academy, centred physically and intellectually in Russia, did the job instead. But that has carried little weight with the independently minded scientists who have played important parts in the counterrevolution of the past five years. They rightly resent the all-union academy's willing administration of the petty and not-so-petty bureaucracy of the Stalin and Brezhnev periods, its persistent cult of the big-shots (not all of whom have been competent) who have been given power over others' lives and work, and its incompetence.

Nobody should be surprised if the academy Yeltsin acquires is transformed by the coming elections into one that is barely recognizable as the present. Whatever happens, it is plain that there must also be a structural shake-up. As things have been, the academy has had three functions. It has been an academy in the usual sense, electing distinguished people to its membership (and arranging for them to be paid extra, and to enjoy the perks of public office). It has been a source of advice to government — by demonstration, not conspicuously good advice. And it has managed directly a substantial and the most important part of the Soviet Union's research enterprise. The first two functions make sense, but the management of a rounded programme of research, both basic and applied, employing some 200,000 people (most of whom must also be housed) has long stretched the academy beyond the limits of its competence. That function must be taken over by others.

Although most of the academy's research institutes are located within Russia, there are enough of them elsewhere (at Kiev in the Ukraine, for example) for people to notice. And, more seriously, nobody can yet know on what scale the republics, either collectively or separately, will in future be able to support research. It is not merely that the structure of the all-union academy is replicated (on a smaller scale) in the other republics, but that there are comparable arrangements for the medical and agricultural sciences, while the quondam ministries of the quondam central government also have (had) their own research establishments. There could well have been two million people ostensibly working at research a few years ago. Almost certainly, the research force will have to shrink, perhaps drastically. That will be a painful process. There will also have to be more subtle structural changes. Links between basic research institutes and the universities are tenuous even when geography is not an obstacle. They will have to be strengthened in the interests of higher education and research. Links between applied research and industrial enterprises are even less substantial. The production ministries' habit of foisting untested innovations on uncomprehending and even resentful factories may account for much of the appalling performance of Soviet industry.

If, collectively or separately, the republics are now to embrace an economic system in which the prices and the supply of various goods are determined by market forces, it is unthinkable that successful industrial enterprises can avoid managing their own research. The danger, in present circumstances, is that the research enterprise will be made to shrink before anybody can take a view of which parts can contribute to a brighter economic future. There is an urgent need for an inventory (and an evaluation) of all the Soviet Union's research, and then a re-assignment of those parts considered worthwhile to the universities or industry. Here is an opportunity at which the World Bank, or the new London-based European Development Bank, should jump. And quickly.

There would, of course, remain a group of research institutes of such size and distinction that they could not be dealt with by forging either academic or industrial links — two molecular biology laboratories in Moscow, the Protein Research Institute at Puschino and the electron-accelerator laboratory at Novosibirsk among many others. What would happen to them? The temptation to let a



newborn academy hang onto them should be resisted. What the new set-up needs, preferably at the level of the centre, is what every other industrial society has — an objective merit-based mechanism for making grants of money for research. It is asking too much of an academy whose members are likely to be the chief beneficiaries of what money there is to carry out that task impartially. Nor, as the experience of the past few decades has shown, can resources be shared between productive laboratories without the malign intervention of power-brokers. These laboratories should also go to a new grant-making agency.

What would then be left for the new academy to do? Plenty. As elsewhere, an academy designed to seek out and acknowledge excellence can have a powerful influence on the standards of the research enterprise. To be effective, it must be independent of government, not its agent. Such an academy is also the natural focus for relations with scientific communities elsewhere: there will be a growing need for foreign travel and placement in the years ahead. But the advice of a genuinely independent academy could be especially influential and valuable just now. Russia and the rest of the old union were rightly proud of their accomplishments in science — a triumph of immense talent over appalling difficulties. Most of those who have made the running on economic reform in the past five years have been members of the old academy, and may yet save the old union from catastrophe. Should not the academy stir itself to save a little of its natural science as well? □

Dead end tunnel

The British government has found a clever way of making sure the Channel Tunnel will make no difference.

BRITAIN, forever boasting of its faithful if reluctant compliance with the European Communities' legislation, is furtively but ingeniously planning the most serious crime in the economic lawbooks — an elaborate non-tariff restraint on trade that will entirely out-class the French ploy, some years ago, of sending imported Japanese video-recorders for inspection to a railway siding at Poitiers. That, at least, is the only rational explanation of the arrangements being made (or not made) to connect the British landfall of the Channel Tunnel, due to be opened in 1993, with the rest of the British Isles. Somebody has worked out that there will be no surge in British imports from the mainland if goods sent for sale in Britain are left to rot in Kent, in the southeast corner of England. But Machiavellian to the end, the British government has cleverly disguised its intentions beneath such a laughable facade of muddle that even the French have not yet seen through the trick.

The plot began a decade ago, when the Thatcher government sponsored legislation to allow the tunnel, but on the strict condition that no taxpayers' funds were used either for construction or in arrangements to handle the extra traffic. So far, the British Government is committed

only to widening a stretch of the only motorway to the Channel coast which is even now a bottleneck, and whose connection with the rest of Britain is the London Orbital Road, the most congested highway in Europe. But the Channel Tunnel is a railway tunnel (some of whose trains will carry motor-cars) so that rail links with the rest of Britain are all-important. That is where the government's fiendish ingenuity has been concentrated.

First, the nationalized railway, British Rail, was asked to devise a route for a high-speed railway from the Channel towards London, whereupon it became plain that the route would pass through many electorally marginal constituencies whose incumbent MPs support the government. They were quickly up in arms at the environmental nuisance the railway would inflict on their electors; the government decreed that much of the railway should be hidden in tunnels, increasing the estimated cost by more than a half, to £3,600 million. The idea was that the funds should be raised privately and the railway operated separately from British Rail, which would have been a minor partner in the enterprise.

The government would have known, of course, that such an investment would not happen spontaneously, especially in a recession, but to make assurance doubly sure, it has for years been musing aloud about its hopes for privatizing British Rail, sowing uncertainty in the minds of railway buffs with money enough to fancy playing with full-size trains. To throw off the suspicion that it might not be serious, the government hit on the clever trick of letting BR spend £160 million buying up houses in suburban south-east London, where its original plan called for a station.

That will not now be needed. In a master-stroke of prevarication, the government has decreed that, if there is ever a high-speed railway, it will follow a different route, crossing to the north bank of the Thames well downstream of London, eventually linking with the rest of the rail network at a grand terminus not yet built (but whose environmental impact, the European Commission complained last week, has been ill-considered). Now the only certainty is further uncertainty. Even if the funds materialize to build the railway, it cannot be finished until a full decade after the opening of the tunnel. During that interval, goods and passengers reaching Britain from the mainland will be fed into a domestic transport system already on the edge of gridlock.

Taken separately, these inconsistencies might be dismissed as incompetence or indecision, which is the cleverness of the plot. Only in combination is their grand design apparent: to keep out imports from the mainland, and all but the most indefatigable of those who live there. Those who believe that what seems to be sheer muddle will make Britain the laughing-stock of Europe should therefore think again, perhaps recalling that when a channel tunnel was first mooted, the British were most of all fearful that it would simply be a way of letting Napoleon's armies walk across. Outwardly, of course, much has changed since then. But who can say that the old xenophobia has entirely been exorcised? □

US makes plea for Japanese SSC help

■ Answer: "no", and an emphatic news conference
 ■ Bush expected to try one last time next month

Tokyo

JAPANESE officials have said "no" as clearly as they know how to US requests for a major contribution to the construction costs of the US Superconducting Super Collider (SSC). The rejection — Japan's first official answer to the United States — came last week at a high-level meeting in Tokyo between US presidential science advisor D. Allan Bromley and Akiko Santo, director general of the Science and Technology Agency. But the door is not shut completely. There is still a possibility that Japan will decide to contribute when US President George Bush meets with Japan's soon-to-be-elected new prime minister in Tokyo next month.

Japanese financial support is vital to the future of the SSC. The US law that established the SSC project requires that at least one third of the more than \$8,000 million needed to build the collider must come from outside the US Federal government. Without a substantial Japanese contribution, this requirement will be very difficult, if not impossible, to meet.

And so a steady stream of top US officials have been visiting Tokyo to court the Japanese government. First to arrive in June last year was a large delegation of US scientists and representative of the Department of Energy (DOE) who called on all of Japan's science-related ministries and agencies. And US President George Bush has repeatedly asked Japanese Prime Minister Toshiki Kaifu to support the SSC whenever the two meet.

But the Japanese response has been decidedly cool. At the last US-Japan summit in July, Kaifu hinted to Bush that it

would be very difficult for Japan to contribute to the SSC because the Japanese government's first priority at present is to renovate Japan's rundown university research system (*Nature* 352, 177, 18 July, 1991).

Japanese government officials cited very similar reasons in rejecting US requests for SSC support at last week's meeting. Santo said at a press conference after the two-day meeting that the Japanese side explained to their US counterparts that contribution to the SSC would cause "difficulties" because of budgetary constraints on government spending, and because of Japan's need to strengthen its basic research system, particularly in the universities.

When asked at the press conference to clarify if the Japanese explanation meant "no" to the US request, Santo confirmed that it did — a view which is echoed by officials of the Ministry of Foreign Affairs, who also participated in the meeting.

The rejection came despite last-minute offers from the US side that were clearly intended to sway the Japanese. At a press conference in Tokyo the day before last week's meeting, Bromley urged Japan's participation in the SSC project as "a full partner" and offered Japan's scientists a bigger role in managing the project (insufficient Japanese input is one of the key concerns of Japanese critics of the SSC).

On top of this, he announced that the United States will probably contribute financially to Japan's Human Frontier Science Program, something Japanese government officials have long been hoping for (see below). He also hinted at possible US support for the Intelligent Manufacturing System project proposed by Japan's

Ministry of International Trade and Industry, as well as the sixth generation computer project proposed by the same ministry.

All of this probably came too late to alter the Japanese position, which has been months in the making. But the Japanese veto is not final. It only represents rejection at the ministry level. A reversal of the decision at the "very top level" is still on the cards, a foreign ministry official says.

At the meeting it was agreed to leave final discussion of the SSC to the US-Japan summit meeting in Tokyo next month between Bush and Japan's new prime minister, who will be elected at the end of this month. Kiichi Miyazawa, a former minister of finance and fluent English speaker, is widely expected to get the job, and it is possible that his new administration will decide to support the SSC so as to start off on a friendly footing with the US administration.

David Swinbanks

INDIA

Skinny cows no threat

New Delhi

US STUDIES on the amount of methane — a potent contributor to the greenhouse effect — produced by Indian cows overstated their threat by a third because they assumed the Indian animals were as large as their US counterparts, a new Indian study finds. Researchers at the Council of Scientific and Industrial Research (CSIR), as part of an exercise to collect data for the 1992 world climate conference in Brazil, say that US researchers simply extrapolated US bovine measurements made "under very different conditions" from those



in India. Indian cows eat and weigh less than US cows. Similarly, methane emissions from Indian rice paddies are less than from US paddies of the same size because of smaller biomass of Indian rice plants. Based on field measurements, CSIR estimates that total Indian methane output is just over 10 million tonnes a year — less than a quarter of the 48 million tonnes a year that the US Environmental Protection Agency and Oak Ridge National Laboratory have calculated. K.S. Jayaraman

A minnow to catch a mackerel?

Tokyo

THE US offer of financial support for Japan's Human Frontier Science Program (see above) is the first official indication that the United States is prepared to put money into the Japanese life sciences programme. A US contribution would provide a substantial political as well as financial boost for the effort's international research on the brain and molecular biology.

Until now the United States has only offered "in kind" support for Frontiers by agreeing to pay the salary for one postdoctoral Frontier fellow to work at the US National Institutes of Health. Even this tiny contribution (about

\$40,000 a year out of a total Frontier budget of \$30 million) is seen as politically very significant by the Japanese, who have had a hard time winning US support for the programme.

D. Allan Bromley, science advisor to the US president, who made the offer last week, did not mention a dollar figure. But an aide was quoted in the press as mentioning "several million dollars". Japanese government officials say they have heard similar rumours but they do not expect the United States to make a firm offer until a decision is made on the SSC, to which the Frontier offer is clearly linked.

David Swinbanks

Evolution of a gadfly

Washington

THERE was something distinctly wrong with the picture: Bernadine Healy, the director of the US National Institutes of Health (NIH), was testifying last week before a congressional committee about the dangers of misusing genetic information. Beside her and echoing her concerns was Nobel Laureate James Watson, director of the US human genome project. Gene therapy pioneer W. French Anderson was nodding agreement on the right.

And sitting just behind them, having orchestrated the whole thing, was Jeremy Rifkin — biotechnology heretic, sworn enemy of genetic release, filer of countless lawsuits, general science pest.

Only a few years ago, prominent scientists were refusing to appear on the same panel as Rifkin, were calling him a "modern-day Luddite", a "fearmonger", a "nut". Now they support his legislation, endorse many of his concerns, sometimes even — hard as it is to believe — talk to him.

Have researchers and policy-makers in US science finally given in to pressure from a self-described radical? Or is it Rifkin himself who has gone soft?

The answer, not surprisingly, turns out to be a little bit of both. Fifteen years of Rifkin's lawsuits, petitions, legislation, press-conferences and general harassment has finally made a dent on the gene scien-

tists. "He's an activist — he's trying to push a point of view that biotech is rushing forward at a pace too fast for society," says Anderson. "But I agree with his ends, if not his means. As always, he can't resist taking potshots at NIH in public, but in private he is much more reasonable. He's



Heretic no more?

trying very hard to develop a consensus and a basis for cooperation."

And, indeed, Rifkin is showing a new desire to work with — if not in — the system. He corresponded with Watson for a year before Watson agreed to support, in principle, a genetic privacy bill that Rifkin drafted and Representative John Conyers (Democrat, Michigan) introduced last year. "It is not unimportant that we got those people in that room [last week]," Rifkin says. "They didn't have to come. This is our first common ground after 15 years of adversity."

But politics speaks louder than philosophy, and last week's extraordinary scene may owe more to Rifkin's congressional acumen than any great shifting of the scientific status quo. If Healy, Watson and David Galas, the head of the genome project at the Department of Energy, were testifying as if their funding depended on it, that is largely because much of it does. In part because of Rifkin's lobbying, Congress has inserted language into the NIH and DOE genome budgets that requires the agencies to spend at least four per cent of their allocations on research on the social and ethical issues of genetic manipulation. Whether or not the officials personally feel as strongly as Rifkin does about the dangers of genetic information, they are being paid to take them seriously.

Nevertheless, it is not insignificant that the genome project's policy-makers were saying the sort of things last week that

Global warming meets genetic engineering

Washington

THERE is really no plausible connection between global warming and genetic engineering. But after about half an hour with Jeremy Rifkin one is not so sure. Rifkin, the president of the anti-biotechnology Foundation on Economic Trends, also happens to be the president of the Greenhouse Crisis Foundation, a three-year old organization that shares the same staff and modest Washington office. Critics have derided him as an opportunist who has made a career of jumping on bandwagons. Certainly his groups' cryptic names do nothing to suggest any special issue loyalty, nor, for that matter, does his entry into global warming activism at the peak of its popularity.

But Rifkin, if nothing else, makes great intellectual arguments. Global warming, he says, would raise the temperature and decrease the rainfall in many agricultural regions. Keeping those areas productive means finding plants that can handle the harsh new conditions. "We won't be able to wait for evolution", he says. "We're going to need genetic engineering." Because he

is generally against the production and release of genetically engineered plants (and animals), it is in his interest to try to prevent an environmental situation that demands them. Q.E.D.

These are some of the other current issues that have drawn Rifkin's attention — and lawsuits:

- **Genetic information:** The Rifkin-drafted Human Genome Privacy Act, introduced in the House of Representatives last year, would forbid government agencies and their contractors from disclosing any individual's genetic information without written consent, with the exception of medical emergencies and criminal investigations. Notably, it does not address what the agency does with the data internally. Rifkin expects the bill pass the House this session, although there is as yet no matching legislation in the Senate.

- **Bovine and human growth hormones:** Despite government studies to the contrary, Rifkin believes that growth hormones have not been proved safe, and may have harmful side-effects, such as 'cow burn-out' and economic damage to small dairy farmers. His group coordinated a lobbying, legal and grassroots campaign last

year that resulted in bans on bovine growth hormone in the dairy states of Wisconsin and Minnesota.

- **Animal patenting:** Patenting animals makes them legally "indistinguishable from microwave ovens," Rifkin says. His group is lobbying in the United States and Europe to prevent patenting of genetically-modified animals. Although the United States granted one patent (for the 'Harvard mouse') three years ago, and the European Patent Office has announced its intention to do the same (see *Nature* 353, 589; 17 October 1991), Rifkin points out that no subsequent US patents have been granted, despite more than a hundred applications in the queue.

- **Genetic release:** Since his first success in delaying government-approved plans to test a genetically engineered microbe that protects plants from frost, Rifkin has fought for increased precautions in releasing genetically modified organisms into the environment.

- **Global Warming:** The Greenhouse Crisis Foundation organizes public education programmes and lobbies for increased research and legislation for safe energy and renewable agriculture. **C.A.**

Rifkin has been pilloried for in the past. "Like all powerful tools, genetic information can be misused and abused. Discrimination based on genotype must be prohibited as a matter of basic civil rights," Healy told the legislators. Now is the right time to be studying and anticipating such problems, she said. "We don't want to have the genome in front of us and then scratch our heads about what this all means."

"The idea that there will be a huge databank of genetic information on millions of people is repulsive," Watson added. He pointed out that funding for research on the ethical, social and legal issues of the genome project have increased every year, from three per cent in 1990 to an estimated five per cent in 1992. "Five years from now, we may be spending ten per cent," he said.

This is heady stuff for Rifkin. "For Healy to sit there and say we've got problems with eugenics and genetic discrimination, I almost feel like I'm being left behind," he marvels. "All these people we've been fighting for the last 15 years had refused to accept the framework [of ethical and social concerns]. Now they talk of nothing else. We've got the molecular biologists accepting that you can't talk about the human potential without talking about eugenics."

Rifkin is not, however, about to pack up and go away. There are still plenty of genetic issues on which he differs from much of the scientific community, including animal patenting and the release of genetically engineered organisms (see below). And even on genetic privacy, Rifkin is not planning to rest on philosophical support from science's movers and shakers — he wants active lobbying.

"If Congress does not pass legislation, we will aggressively oppose the human genome project," he says. "We can make a hell of a stink. I believe that there are enough scientists out there to support us. I'm not naive — they're not going to man the barricades. But as of [last week's hearing] there is a new situation."

Perhaps. But 15 years of bad blood takes a while to get over. Anderson recalls how he gave Rifkin a list of researchers to contact for support on the genetic privacy issue. "Most people wouldn't even talk to him," he says. "Rifkin still feels so strongly on some issues. If someone comes up with a germline gene therapy [manipulating genes in reproductive cells], he'll file a lawsuit as fast as his fingers can fly over the word-processor keys."

Indeed, lawsuits are no way to make friends. But Rifkin believes that researchers need people like him. "When you're on the cutting edge of science," he says, "you don't stop to be reflective. You need pressure from the outside." With a little congressional prodding, genome researchers, at least, seem to agree.

Christopher Anderson

Privacy bill vetoed

CALIFORNIA governor Pete Wilson last week vetoed what would have been the first US law protecting the privacy of genetic information. Although the bill, which was drafted by Paul Billings, a geneticist at the California Pacific Medical Center, had passed both houses of the state legislature, Wilson argued that the issues surrounding genetic information needed more airing. "Since this field [of genetic diagnosis] is so new and so rapidly expanding," he wrote in a 14 October letter to legislators, "I am concerned that we are providing a remedy for a problem whose nature and magnitude are not yet sufficiently defined." The bill would have placed an eight-year ban on the use of genetic information in insurance and employment decisions (see *Nature* 353; 5 September 1991). C.A.

OSI confirms mix-up

AN independent test by the US National Institutes of Health Office of Scientific Integrity (OSI) has confirmed that the AIDS blood test developed by researcher Robert Gallo was based on a French virus sent to him by the Paris-based Pasteur Institute. This appears to confirm what both Gallo and Pasteur researcher Luc Montagnier have already determined through genetic analysis of their own viral samples from the 1984 period in which the AIDS virus was first isolated. As part of its continuing investigation of Gallo's laboratory, OSI in July asked Roche Diagnostics Research to conduct an independent analysis of the samples. According to a terse OSI statement released last week, Roche's DNA sequencing studies showed "no evidence that the French [AIDS virus] isolate was misappropriated by Dr. Gallo or his staff in preparing the materials for the blood test."

Further, the Roche study confirmed what is now widely known—that a virus from the French patient LAI contaminated cultures in laboratories in France, England, and the United States.

Finally, OSI points out that it "has determined and announced previously that since Dr. Gallo had several HIV isolates in his laboratory, as well as a continuous culture of a virus not related to the French virus, he had no need to appropriate the virus from the Pasteur Institute." The *Chicago Tribune*, in an article this week, quotes former OSI deputy director Suzanne Hadley as calling that statement "completely erroneous". "We never made such an announcement," she is quoted as saying. But in a 5 October, 1990 statement, then-acting NIH director William Raub announced that OSI "has concluded that Dr. Gallo had a substantial number of HIV detections and isolations from several different sources at the critical time that [the French virus] was being grown..." C.A.

PETA wins one

PEOPLE for the Ethical Treatment of Animals (PETA) last week won a retraction from a magazine that had published an article alleging that the US animal-rights group was financially corrupt and that its best-known publicity photos were staged (see *Nature* 343, 580; 1990). As part of an out-of-court settlement, *Washingtonian* magazine agreed to print in its next issue a lengthy "correction and clarification" that withdraws most of the more damning allegations in a critical 1990 article by freelance journalist Katie McCabe. Among the allegations that McCabe retracts is that PETA co-founder Alex Pacheco had tied one of the 'Silver Spring Monkeys' in an unfamiliar experimental apparatus to take a picture that has since become one of the symbols of the animal-rights movement. The monkey was placed in the apparatus by laboratory personnel as part of a scheduled experiment, McCabe concedes. *Washingtonian* also retracts allegations that PETA had engaged in secret accounting procedures to channel funds to other animal-rights groups and activists; the allegations were based solely on interviews with former PETA employees who had been dismissed, one of whom subsequently retracted his statements after PETA sued him. C.A.

First to market

Sydney

AUSTRALIA is the first country to allow the sale of a genetically altered living organism for general commercial use. An Australian company, Biocare Technology Pty Ltd, is now marketing a pesticide that consists of a slightly altered version of a natural biological counter to the plant disease Crown Gall.

The pesticide has been used in one Australian state for three years under the name Nogall. But Allen Kerr of the University of Adelaide, who led the team that discovered Nogall, said that this was the first time that a genetically altered organism had been approved by any country for unrestricted national sale.

Kerr's team discovered the original Nogall organism, a strain known as K-84, during a lengthy search for a biological control for Crown Gall. Trials showed that after a period of exposure to Crown Gall, K-84 lost its effectiveness because it transferred a plasmid to the virus. The plasmid DNA allowed the virus to generate an antibody against the attacking K-84. That problem was overcome by altering the genes of K-84 to prevent the transfer.

The state of New South Wales approved the pesticide in 1988, but the Federal Government only granted registration last month. Biocare Technology has applied to register Nogall in the United States and Japan.

Mark Lawson

Canada back in the fold?

FOUR months after the plans of Canadian astronomers to collaborate with their international colleagues in building two 8-metre optical/infrared telescopes seemed to have evaporated, some hope has re-emerged.

Earlier this year, the Canadian National Research Council (NRC) said that it could not afford to meet its planned 25 per cent contribution to the \$176-million international Gemini project, which will build twin 8-metre telescopes in Hawaii and Chile (see *Nature* 351, 680; 27 June 1991). At the time, astronomers in the United States and Britain—the other two partners in the effort—assumed that Canada was out of the project for good.

But after the first meeting of the Gemini project's managing board — held in Ottawa earlier this month, with Canadian officials present — project leaders are now hopeful that Canada will, after all, join the project. The solution under discussion is for Canada's Natural Sciences and Engineering Research Council to join with NRC in funding the project. With some additional money from a consor-

tium of Canadian universities, Canada may be able to take a 15 per cent stake in Gemini.

Don Morton, from NRC's Herzberg Institute of Astrophysics, who in June had to write to Canadian astronomers to break the bad news about the Gemini project, is now talking of a possible "reduced, but significant contribution", although he declines to discuss any precise figures. Project scientist Gordon Walker, from the University of British Columbia, believes NRC is responding in part to strenuous lobbying by Canada's astronomers, in particular by postdoctoral and graduate researchers, urging NRC to think again.

Continued Canadian involvement in the project, even at a reduced level, is important to Britain and the United States. British spending on Gemini has already been delayed by two years, and the UK Science and Engineering Research Council is considering a plan to cut several million pounds from the British contribution over the first four years of the project, and to spend this money instead when the telescopes are nearing completion.

Peter Aldhous

OZONE DEPLETION

Montreal Protocol still too lax?

London

EVEN if all nations abide by international controls on ozone-depleting chemicals, ozone loss in mid-latitudes over the next decade may match that experienced in the Antarctic during the 1980s. This week, a team of leading atmospheric scientists asked by the United Nations Environment Programme (UNEP) to review the damage to the Earth's stratospheric ozone layer warned that reducing these ozone losses will require tighter limits on the emissions of chlorine-containing compounds than those contained in the Montreal Protocol, amended in London only last year.

The review team's assessment assumes that the amended Montreal Protocol will be fully obeyed, although too few countries have so far ratified the amended protocol for it to come into force (the United States and many European countries are among the holdouts). India, which has the potential to be a huge emitter of ozone-destroying chemicals, says that it will only join the protocol once the London amendments have come into effect.

Under the amended Montreal Protocol, the loading of chlorine in the stratosphere will peak at 4.1 parts per billion by volume (ppbv) at the turn of the century, and will not return to 2 ppbv — the concentration at which the Antarctic ozone hole appeared — until 2060. According to the UNEP review, if chlorofluorocarbons (CFCs), carbon tetrachloride and methyl

chloroform are phased out by 1997, chlorine loading would peak at below 3.9 ppbv, and the atmosphere should return to pre-ozone hole conditions by 2053. (The amended Montreal Protocol calls industrial countries to abandon CFCs and carbon tetrachloride by 2000 and to cease using methyl chloroform by 2005.)

The new UNEP assessment, together with the discovery that this year's Antarctic ozone hole is the worst on record, will increase pressure to tighten the Montreal Protocol, when it is reviewed next October. By then, policy-makers should also have the results of a new £14 million international experiment to study Arctic ozone depletion. The European Arctic Stratospheric Ozone Experiment starts next month, and will involve some 250 researchers from 17 countries.

The aim of the new experiment is to study this winter's Arctic polar vortex, monitoring ozone destruction through the evolution of a single air mass. Researchers hope that these data will convince policymakers that Northern Hemisphere ozone depletion is a serious problem. (Ozone levels north of 40 degrees N are thought to have declined by 5 per cent between 1970 and 1986.) The experiment, which runs until March 1992, will involve ground-based instruments, research balloons and two aircraft, and is funded by national governments and the European Commission.

Peter Aldhous

European gravity detector in peril

London

WHILE US astronomers toast the recent decision of Congress to allow the construction of two gravity-wave observatories in the United States, their colleagues across the Atlantic face the loss of a planned matching detector. The UK Science and Engineering Research Council (SERC) is poised to cancel its proposed £5.5 million contribution to a £40 million German observatory, a move that would jeopardize the entire project.

Gravity waves are 'ripples' in space-time, which the theory of general relativity predicts should result from the movement of huge masses in space — the collapse of a star's core during a supernova, for example. The Anglo-German gravity-wave detector was one of several projects put under review when SERC ran into financial problems a year ago. Asked to plan its future spending within a strictly limited budget, SERC's ground-based programme committee has now decided that it cannot earmark money for the project.

The observatory's fate is not yet sealed — the decision on the British contribution rests ultimately with SERC's governing council — but the project's supporters are pessimistic. SERC chairman Sir Mark Richmond's public comments on the project have been negative, describing the observatory as an expensive machine designed to perform a single experiment for a small group of researchers (only several dozen British astronomers were expected to be actively involved). Since taking over at SERC, Richmond has made it clear that he wants to reduce SERC's spending on large facilities, except where these are used by a large community of researchers.

If SERC does back out of the gravity-wave detector, the German research ministry (BMFT) may abandon the project, say officials in Bonn. BMFT faces its own financial difficulties, with the high costs of German unification affecting all areas of government spending, and is looking for ways to trim its budget.

The loss of the German observatory would leave European hopes for a gravity wave detector pinned on a proposed Franco-Italian collaboration, also yet to receive a firm guarantee of funding. Jim Hough, from the University of Glasgow, and the leading British researcher for the Anglo-German detector, argues that the presence of two European observatories in addition to the US detectors would increase the likelihood of detecting waves from background noise, and help to locate the direction in which they are travelling.

Peter Aldhous

NOBEL PRIZE: PHYSICS

A gift for unraveling complexity

London

ALTHOUGH Pierre-Gilles de Gennes of the College de France won this year's Nobel prize in physics for his work on polymers and liquid crystals, the decision reflects just as much his contribution throughout condensed matter physics. His work has been instrumental in defining the topic



Father of the LCD display

itself — that is, in demonstrating that principles can be identified which describe the behaviour of condensed systems as diverse as magnets, superconductors and assemblies of macromolecules.

De Gennes is known for his ability to see the underlying physics in systems of seemingly intractable complexity. In this sense, his work follows the spirit of that of the Soviet physicist Lev Landau, who showed that phase transitions in disparate media commonly share universal features.

De Gennes' early work on phase transitions involved the "classical" problem of ordering phenomena in solid-state magnets, and he contributed further to the understanding of "hard" condensed systems with work on superconductivity. But it is his research into soft condensed phases that has had the greatest impact, particularly those — such as polymers and liquid crystals — for which the generic term "complex fluids" has now become fashionable. De Gennes' most telling insight lay in the proposal that concepts developed to explain classical phase transitions could be carried over to these considerably less tidy systems.

Working at the University of Orsay, de Gennes showed that transitions between

ordered phases of liquid crystals bore analogies to the transition from a normal to a superconducting state. In transforming from a so-called cholesteric to a smectic phase, liquid-crystal molecules initially aligned in a twisted manner lose this twist and become ordered into a parallel array.

Transitions in liquid crystals have at least the simplifying characteristic that the molecules tend to align collectively into ordered structures. Polymers rarely show such cooperatively but instead display a disordered tangle of chains. Yet de Gennes succeeded in extending to polymer solutions the ideas developed to describe phase transitions in much "cleaner" systems. He showed that the physical properties of polymers obey "scaling laws." Scaling is a central tenet of the modern theory of phase transitions.

In surface science, de Gennes advanced the understanding of wetting dynamics by providing a theory for how liquid droplets spread on smooth and rough substrates. His work on the adsorption of polymers at solid surfaces has now led him to display an interest in applied fields such as tribology and adhesion — traditionally engineering disciplines, to which de Gennes will no doubt show that condensed matter physics has a contribution to make.

Philip Ball

CHEMISTRY

Bringing NMR to the masses

London

COMPLETING a clean sweep by European researchers of this year's science Nobel Prizes, the 1991 award for chemistry has gone to Richard Ernst of the Swiss Federal Institute of Technology in Zürich. Ernst is responsible for two refinements that helped to transform nuclear magnetic resonance spectroscopy (NMR) from a specialist tool into a staple of modern chemistry.

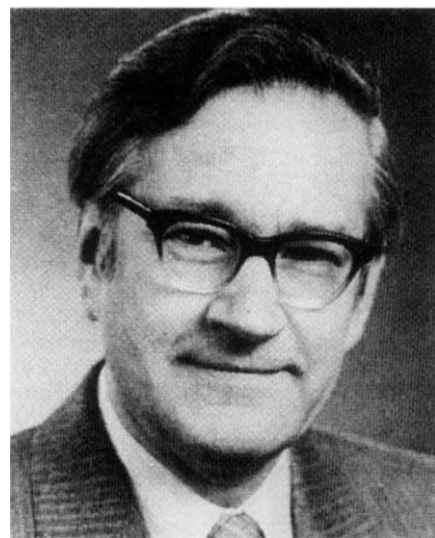
Before Ernst's work, NMR involved applying a slow sweep of varying radio frequencies to a chemical sample held in a magnetic field. At particular frequencies, the spins of many atomic nuclei reorientate themselves with respect to the magnetic field, which induces an electric signal in the detectors of an NMR spectrometer. Because these 'resonance frequencies' vary according to the chemical environment in which the nuclei find themselves, the pioneers of the technique realized that NMR could be used in chemical analysis. Early NMR spectrometers, however, lacked sensitivity. Only a handful of nuclei — principally hydrogen, fluorine and phosphorus — yielded signals strong enough to distinguish their resonance frequencies reliably.

In the mid-1960s, Ernst dramatically increased the sensitivity of NMR by replacing the gentle sweep of varying fre-

quencies with short, intense radio pulses. Ken Packer, from British Petroleum's Sunbury Research Centre, likens the approach to "hitting a piano with a hammer", rather than pressing each key in sequence. By applying a mathematical operation called a Fourier transformation to the complex signal given off by the nuclei, Ernst was able to reconstruct standard NMR spectra.

The Fourier transformation method increased NMR's sensitivity by more than tenfold, greatly expanding the number of different nuclei that could be studied. Most important, chemists could for the first time reliably determine the resonance frequency of carbon-13 nuclei, which yield useful NMR spectra (unlike the more common carbon-12) but which were present in insufficient quantities in chemical samples to be analysed successfully using the first NMR spectrometers.

Ernst's second major contribution was to pioneer two-dimensional NMR. Instead of using single radio-frequency pulses, Ernst and his colleagues began to experiment with sequences of pulses. In two-dimensional NMR, samples are repeatedly subjected to a pair of pulses in close succession, and the interval between the pulses in each pair is varied. The resulting signals can then be measured both as a function of time during each detection (as



Explored the mathematics of resonance

in Fourier transform NMR), and as a function of the interval between the two pulses in each pair. These data then give two-dimensional spectra that can show which nuclei in a molecule lie close together.

This new development allowed NMR to be applied to the study of larger molecules than was previously possible. The result was that NMR became a key tool in biochemistry, launching a new range of structural studies of biological molecules in solution, and even of their interactions with drugs and metal ions. **Peter Aldhous**

Chernobyl fire chronology

Moscow

ON 11 October at 8:09 p.m., 20 minutes after its No. 2 power unit's fourth turbo-generator had been shut down for scheduled repairs, fire once again swept through the Chernobyl Nuclear Power Plant (NPP). Operators had accidentally connected the generator to the plant's power network, forcing powerful currents through its wiring and igniting its insulation.

Sixty seconds later, at 8:10 p.m., the NPP's operator pushed the emergency button in a bid to shut off the No. 2 power unit's reactor. Oil was vented from the turbo-generator and operators began substituting hydrogen inside the reactor with nitrogen, in a fruitless attempt to check the spread of the fire. Special seals leading into the generator had apparently broken when the insulation caught fire. As a result, hydrogen escaped from the unit, subsequently catching fire and exploding. The engine room's roof started burning, and, according to the late reports, four roof girders caved in.

Automatic fire extinguishers then activated, and fire-fighters arrived at the scene some time later, putting out the fire by

11:30 p.m. The No. 2 power unit's reactor is now being cooled off.

Experts voiced reassuring statements during a news conference in Kiev last week, saying that there were no fatalities and that no radiation had spewed into the atmosphere. Only the engine room proper has been damaged, they added. However, the accident has entailed considerable losses, including the halting of the damaged power unit. The precise causes of the accident are now being verified.

Curiously enough, the *Trud* newspaper carried an interview with Victor Bryukhanov, who headed the Chernobyl NPP prior to the April 1986 disaster, on September 17. It was headlined "The Chernobyl NPP Must Be Stopped Completely". This was Bryukhanov's first interview upon his release from prison. He was been sentenced to 10 years in prison during the Chernobyl trial, but has just returned home after serving half of his term in a medium-security labour camp. Until earlier this year, he was engaged in corrective labour at a place not far from Kiev that can be reached by car in three hours' time.

Yuri Kanin

SOVIET SCIENCE

Russia takes over academy

Moscow

IN science, like nearly every other aspect of Soviet life, last August's attempted coup has become a catalyst for radical change and disintegration. As institutions fall, even the country's supreme scientific organization — the USSR Academy of Sciences — has not remained immune.

At the Academy's general meeting last month, academy members voted 160 to ten to "return to the USSR Academy of Sciences the status of the Russian Academy of Science." This will put the Academy and its research institutes under Russian jurisdiction, although just what the Academy will consist of in the future is an open question. Russia has just created its own Academy of Sciences, and is unlikely to split its scarce funding two ways. Another major Soviet scientific structure, the USSR State Committee for Science and Technology, has also transfer to Russia. According to Vladimir Shoron, chairman of the Russian Supreme Soviet Committee for Science and Public Education, the Russian Council of Ministers decided on 8 October to take the science and technology committee under its wing.

Russia is the largest Soviet republic, both in geographic area and in scientific potential. The majority of the 365 scientific institutions of the USSR Academy, are in Russia, as are 95 per cent of its members. And until recently, Russia was

the only one of the newly-sovereign states which had no academy of its own. So in June 1990 — before anyone suspected that the USSR Academy might soon be put under Russian jurisdiction — the Russian Supreme Soviet decided to establish a Russian Academy. In March 1991 an organizing committee was set up to develop the charter and guidelines for the Academy. The newly-formed regional organizing committees have nominated candidates and will hold first elections for Academy membership in December.

Yet many questions still remain. Now there will be two academies in Russia: One of them, the Russian Academy, has no material base, institutions or laboratories, but has been promised relatively solid financial backing from the Russian government. The other, the now-Russian-controlled USSR Academy, has numerous institutions, but little political support. Its researcher cannot realistically hope for adequate funding. As USSR Academy researchers were told at the general meeting last month, the Academic institutions will have no hard currency to subscribe to foreign scientific publications in 1992. Rank-and-file research associates are paid less than janitors. And researchers are afraid that the USSR Academy will wither under Russian rule, something that is likely to only worsen the brain drain and exodus of gifted young scientists.

Yuri Kanin

Trouble at AFRC

London

AGRICULTURAL research is the latest area of British science to fall victim to the economic recession. Tom Blundell, director-general of the Agricultural and Food Research Council (AFRC), says that his organization faces a £10 million shortfall in the 1992-93 fiscal year. In large part, Blundell blames the collapse of the British property market, which has left AFRC unable to sell several large laboratories that were closed during the cutbacks and restructuring that preoccupied the council during the 1980s. In the context of AFRC's smaller budget (the council receives only about £90 million a year from the Department of Education and Science), AFRC's plight is similar to that which has afflicted the Science and Engineering Research Council over the past year.

Blundell, who first revealed his council's problems in an interview with the *Times Higher Education Supplement*, says that AFRC had planned to receive some £7 million from the sale of its unused buildings and land next year. Most of this money was expected to come from the sale of the Houghton Laboratory in Cambridgeshire, a former base for AFRC's poultry research, or the Hurley Research Station in Berkshire, formerly part of AFRC's Institute of Grassland and Animal Production. Blundell had hoped that one of these laboratories would be bought by a company seeking new headquarters offices. But with most British business concentrating on survival, rather than relocation, no buyers have come forward.

Blundell says that AFRC is also faced with a bill for the pension funds of retired researchers from AFRC's former Institute of Horticultural Research. Although this institute has now been reconstituted as a separate agency, Horticultural Research International, linked to the UK agriculture ministry, an Act of Parliament is needed to transfer the pension fund from AFRC. With the government now concentrating on preparing for a general election, Blundell is concerned that the legislation may be delayed.

Blundell hopes that AFRC's finances can be eased by an infusion of new funds in the government's public expenditure settlement, due next month. But he says that many other public sector agencies also face problems selling their empty buildings, so the Treasury may be reluctant to compensate AFRC for much of the shortfall. If little extra money is found, the bulk of the necessary cuts must come in spending on equipment at AFRC's own institutes and in the council's support for university researchers. The cuts would temporarily reverse the council's long-term plan to increase by 50 per cent its spending in the universities, now running at just under £20 million a year. Peter Aldhous

Fallacious claims for HGP

SIR — Maddox has argued¹ for the value of the Human Genome Project (HGP). Although his case is largely based on the obvious ability of the project to generate enthusiasm in the molecular biology community, he also argues that it will (1) provide important medical benefits in the form of significantly enhanced diagnostic abilities and (2) help understanding of the evolution of the genome itself. These last claims are unwarranted at present².

The argument that medical benefits would be obtained by a comparison between any arbitrary sequence and a "presumably representative" sequence of the human genome is flawed. Divergences would presumably be indicative of genetic abnormality, perhaps even disease. The "representative sequence" would thus be used for diagnostic purposes.

The trouble with this argument is that there simply is no such entity as a "representative sequence" of the human (or any) genome; the amount of variability without loss of function in the DNA sequence of any natural population is far too great. Consider, for example, the DNA sequence of a haemoglobin, perhaps the best characterized of human proteins³; of 2,583 possible base substitutions of the DNA coding for the α - and β -chains of haemoglobin, only 1,690 would result in amino-acid substitutions because of the degeneracy of the code. Of these, only 575 (about 45 per cent) would result in a change that could be detected by electrophoresis. This has already led to the detection of about 450 variants, almost half of which are fully functional and, therefore, physiologically silent.

The true amount of protein polymorphism (not revealed by electrophoresis) without loss of function is, presumably, even higher. DNA polymorphism, the relevant factor when considering genomic sequences, exceeds protein polymorphism by at least another order of magnitude because of random nucleotide polymorphism, the degeneracy of the genetic code and variability of tandem repeat sequences that generate highly variable regions.⁴

Among all these fully functional sequence variants, it is fallacious and even dangerous to call any one "normal", simply because any notion of normalcy that goes beyond functionality has no scientific, medical or philosophical basis. Further, there is simply no known way in which the highly irregular DNA sequence information can be "averaged" to provide a "representative sequence" for storage in a database and used for normative comparison. All that can be done, if a single sequence is to be stored, is to

choose one of the many normal sequences, perhaps the most common. But such a sequence would be of little diagnostic value. A sequence that diverges from it at many different points could yet code for a fully functional protein whereas even a single difference at a critical site might destroy function (as in the case of the sickle-cell haemoglobin). If accurate medical diagnosis is the purpose of the use of sequence information, each arbitrary sequence will have to be independently judged for functionality.

This argument does not even take into account that many of the genes implicated in disease have low or highly variable expressivity⁴. In such circumstances an extragenetic factor is critical to the aetiology of the disease and there is no reason to believe that such an environmental component is in general due to the epistatic effect of genes at other loci. Sequence information is of little value in these cases, since what has to be understood are the interactions with the environment which, at least to a first approximation, consist of, or are mediated by, biological entities at higher levels of organization than the DNA sequence. Once again the quest for DNA sequences offers little direct hope for therapeutic intervention.

These criticisms of the goals of HGP are pertinent only insofar as that project pursues the blind quest for complete sequences. They do not argue against the limited goal of linkage mapping that would locate the relative position of functionally relevant loci on chromosomes. There is little reason to doubt that such a linkage map, and limited sequencing of these and other regions of interest, will provide both medical and evolutionary insights of value. Perhaps this is where the genetics community has shown most wisdom in adopting the two-stage approach to HGP. During the first stage, such linkage maps will be constructed while sequencing techniques are honed. These goals are uncontroversially worthwhile; it is the second stage of complete sequencing (which Maddox correctly identifies as the core of HGP) that remains suspect.

SAHOTRA SARKAR
ALFRED I. TAUBER

*Theoretical Biology Group,
Boston University,
745 Commonwealth Avenue,
Boston, Massachusetts 02215, USA*

1. *Nature* **352**, 11 (1991).
2. Tauber, A. I. & Sarkar, S. *Perspectives in Biology and Medicine* (in the press).
3. Bunn, H. F. & Forget, B. F. *Haemoglobin: Molecular, Genetic and Clinical Aspects*, 381–451 (Saunders, Philadelphia, 1986).
4. Vogel, F. & Motulsky, A. G. *Human Genetics: Problems and Approaches*, 2nd edn, 435–446 (Springer, Berlin, 1986).

Figures in dispute

SIR — Referring to indirect costs at the University of Michigan (*Nature* **353**, 196, 1991) you state: "This is not a dispute about minor errors in a scientific manuscript, but one in which millions of dollars have been lost to scientific research. The disputed charges in Michigan alone add up to \$8.3 million." This statement is false. The charges being disputed are 15 per cent of about \$2 million, and the government has not yet been charged any of this, neither \$300,000 nor \$8.3 million. The issue of allowable costs has arisen during negotiation prior to the setting of overhead rates, not over indirect costs already billed. In the course of these negotiations, government auditors questioned \$8.3 million in university expenditure of which the university proposed to charge government grants and contracts 15 per cent as overhead. The university volunteered to withdraw more than \$6 million, leaving 15 per cent of \$2 million in dispute. The cost of the Rose Bowl tickets is not in that \$2 million.

THOMAS M. DONAHUE
*Department of Atmospheric,
Oceanic and Space Sciences,
Space Research Building,
University of Michigan,
Ann Arbor,
Michigan 48109-2143, USA*

Science in India

SIR — Your report (*Nature* **352**, 465; 1991) of the plan to broadcast a series of 130 radio programmes about evolution on All India Radio (AIR) should have made it clear that the National Council for Science and Technology Communication (which is supported by the Government of India) is an equal partner in the venture with AIR and, in particular, has been responsible for producing the wall charts and kits that will be distributed to registered listeners as the programmes unfold.

NARENDER K. SEHGAL
*National Council for Science
& Technology Communication,
Department of Science & Technology,
Technology Bhawan,
New Mehrauli Road,
New Delhi 110016, India*

Parking fine

SIR — A better solution to the parking problem than that recently proposed by Daedalus (*Nature* **353**, 216; 1991) would be to adopt the method used by supermarket trolleys, giving a new meaning to the word 'hatchback'.

BRENNIG JAMES
*Cherry Orchard,
Marlow Common, Bucks SL7 2QP, UK*

Better answers needed

SIR — The Manifesto for British Science (*Nature* 353, 105; 1991) began by making the important point that it is the way that money is spent, rather than the absolute amount, that is the most important aspect to get right. There were also the important, and we think widely accepted, points that an extension of the period of the first degree, as well as a more structured approach to postgraduate research training, should be considered with some urgency. The proposal that research student stipends and academic salaries should be increased substantially was also welcome, if diverting attention from the main problem and contrary to the opening gambit. However, it was almost with disbelief that we read that a principal answer to the problem should be to double the numbers of research students. Certainly contemporary society demands that young people should receive a good scientific and technical background, and that many should also have some research experience. But, to suggest that the panacea is to double the number of research studentships seems not only foolhardy but also to imagine that young graduates are fools.

In more than 15 years of working in research, we have rarely heard young researchers suggest that their major complaint is one of inadequate personal reward (even though that would be reasonable in many cases). Almost exclusively, the complaints have been because of not knowing if or when the next job will be available. Why would intelligent young graduates want to enter a discipline where there were not enough positions (let alone money) for the numbers already in the race? If the manifesto authors believe that more money alone will tempt potential researchers, why is it so difficult to find good new graduates to fill current, relatively well-paid research assistant posts where they may normally also register for a postgraduate degree? Further, why are there so few applicants for the better paid research fellowships?

Perhaps a more useful way of spending limited amounts of money to encourage young people to take an interest in research, which would complement the desire for four-year science degrees, would be to support sandwich degrees more adequately and get young people into research laboratories as part of their first degree. To persuade them to pursue research further will, however, still require them to observe that research is organized on a rational basis, and a fundamental rethink of careers for researchers. The difficulty here seems to be that the present administrators of academic research are so busy blaming

someone else for the problem that they will not face up to the task of providing an adequate career structure for research scientists. This issue was raised in some detail at the meeting called to discuss the manifesto, although only touched on in a subsequent leading article and News item in *Nature* (353, 195 & 203; 1991). One might hope that the editor and writers of a manifesto for such an august organ would be able to examine the defects within the scientific community as well as those of government. Instead we watch with despair, as they prefer to climb in and out of the crumbling ivory towers inhabited by the present administrators of academe.

STEPHEN J. HOPKINS

*University of Manchester,
Rheumatic Diseases Centre,
Hope Hospital, Salford M6 8HD, UK*

NANCY J. ROTHWELL

*Physiological Sciences,
University of Manchester,
Manchester M13 9PT, UK*

SIR — The Institution of Professionals, Managers and Specialists (IPMS) represents more than 4,000 members working in research councils, and over 15,000 of our other members are scientists working predominantly on research. We welcome the initiative taken by *Nature* in publishing its Manifesto for British Science, and in particular the call for the appointment of a Minister for Science and for the government to end its underfunding of research and we endorse the call for stability in strategy and an end to the switchback of policy changes.

But we believe that the manifesto needs strengthening in two areas and that its tone towards, and policy for, the research councils is fatally flawed.

The manifesto makes much of the need to improve the number and pay of PhD students and the pay of those who supervise their work within the higher education institutions (HEIs). But that will not solve the problem. The relative pay for all scientists needs to improve if we are to attract and retain high-quality students. Sir Mark Richmond stressed the importance of improving pay — he might look to the employees of his own research council where scientists earn less than equivalent engineers. Repeated attempts by IPMS to end this anomaly have been blocked by the councils. So what chance have we of convincing the wider public of the merits of our case?

Data from the 1990 New Earnings Survey show that scientists and mathematicians receive less than two-thirds of the earnings of finance, insurance and tax specialists. The government, as the largest single employer of scientists with

the heads of the research councils, should take a lead in redressing this situation.

An additional problem is the prevalence of short-term appointments, within both the HEIs and research council institutes. As we pointed out in our evidence to the Royal Society: "The increase in short-term appointments has had a disastrous effect on career development, continuity of programmes and stability of employment. The experience of IPMS members is that the system is disruptive, inefficient and unfair."

The proportion of science and technology staff on short-term contracts in universities has risen from 25 to 42 per cent. Within the research councils, up to 80 per cent of new appointments are on a short-term contract basis. There is little prospect of recruiting more scientists unless there is proper career development and greater security of employment. We are not advocating a cosy life for scientists but one in which vigorous science can take place within a properly managed career structure.

The manifesto's proposals for the dismemberment of the research councils would be disastrous. We agree with Dr Jeremy Bray that the absorption of the research institutes by the operational departments would result in no benefit but major disruption of research. Almost 80 years ago, Haldane specifically proposed that the research units be separate from the executive departments so that those beholden to ministers could not direct research programmes nor suppress their results. The research institutes have a crucial role to play in bridging the gap between basic research and its practical application which is still one of the weakest aspects of the UK system.

This antipathy towards research councils and their institutes pervades parts of the manifesto. Like Lord Porter, you believe that the "centralization of research facilities should be reversed". But the view that research can be confined to small teams working in HEI laboratories is outdated. Large facilities are necessary in many areas of work, a view confirmed by many small HEI research teams which make great use of research council facilities, for example the Nuclear Structures Facility at Daresbury. While we accept that the dual-support system is not providing HEI research teams with the resources they need, to attack research council institutes is to target the wrong enemy. Our major problem is a government which changes priorities without warning and keeps researchers, wherever they are located, scandalously short of resources.

VALERIE ELLIS

*Institution of Professionals,
Managers and Specialists,
75–79 York Road, London SE1 7AQ, UK*

Schizophrenia in the melting pot

Julian Leff

The puzzle of schizophrenia has been approached from many directions. But, despite the development of molecular genetic techniques, the origin of the condition remains obscure.

IN many ways, the tenth symposium on the psychotherapy of schizophrenia held on 11–15 August in Stockholm was an extraordinary meeting, mainly because of the focus on the curability of schizophrenia by talk therapy. Most of the speakers were psychoanalysts specializing in the psychotherapy of patients with schizophrenia. This method of treating schizophrenia has few adherents outside Scandinavia, and indeed the meeting was attended by only a small contingent from the United States and the United Kingdom. At one time a psychoanalytic training was almost obligatory for US psychiatrists, many of them taking on schizophrenic patients for psychotherapy. But biological psychiatry is now in the ascendant, as it is in the United Kingdom, and psychoanalysts have been largely ousted from the field of schizophrenia by the introduction of antipsychotic medication and the insistence on its use by health-insurance companies. Although the effectiveness of drug treatment for both the acute treatment and maintenance of schizophrenic patients has been established beyond doubt, the value of any form of psychotherapy remains unproven.

But even more than drug treatment, adherents of the psychotherapy approach saw the advances in psychiatric genetics reported at Stockholm as a major threat. Those with the extreme view that schizophrenia can be cured by psychoanalysis would find their position hard to sustain if faulty genes were responsible for the condition rather than faulty parenting. Three years ago, *Nature* published a paper in which Sherrington *et al.*¹ claimed by linkage studies to have found the location of a 'schizophrenia gene' on chromosome 5. The following year another group claimed to have identified a different locus².

Replication

If these were dependable findings, the meeting in Stockholm might not have taken place, but several subsequent attempts to replicate the results have been unsuccessful. Indeed, at the second international conference on psychiatric genetics, held in London at the same time as the Stockholm meeting, Gurling and Sherrington reported a failure to replicate their original finding. These authors, provoked by the contradictory results obtained by other researchers,

had conducted further linkage studies on the same families that they studied in 1988, using additional markers from the same region of chromosome 5. Disappointingly, they were unable to establish linkage between schizophrenia and any of the new markers, casting doubt on the finding they first reported.

The history of research into schizophrenia is replete with 'breakthroughs' which have later turned out to be illusory. But why should molecular-genetic studies be so fallible in psychiatry when they have yielded such solid discoveries in other branches of medicine? The answer probably lies in the complex nature of severe mental illness.

As a diagnostic entity, schizophrenia is nearly 100 years old, although it was born as dementia praecox and given its modern name by Bleuler in 1911. The condition has elastic boundaries, particularly in the United States where in the 1970s it was twice as extensive as in Britain. At that time, patients who would have been diagnosed with manic or depressive psychosis by British psychiatrists would probably have been diagnosed schizophrenic in the United States. As a consequence, many of these patients received treatment that in Britain would have been judged less than ideal. Following the exposure of these transatlantic differences by epidemiological studies, and a swing away from psychoanalysis in the United States, the definition has now narrowed considerably, according to the diagnostic and statistical manual (IIIR) of the American Psychiatric Association. In fact, fewer patients are now diagnosed schizophrenic in the United States than in Britain. The diagnostic concept remains fluid because there is no pathological test for schizophrenia. Consequently, it is a matter of opinion where the boundaries of the phenotype should be drawn.

Earlier geneticists studying like and unlike twins, and adopted-away children of schizophrenic mothers, introduced the concept of schizophrenia spectrum disorder. This broadened the phenotype drastically to include people with odd personalities but no hint of psychosis. The suspicion is that this strategy was invoked because there were too few cases of 'pure' schizophrenia in the samples studied for legitimate statistical analysis. In the event this may have muddled the waters, given the findings

presented at Stockholm by Pekka Tienari (University of Oulu, Finland). Tienari has repeated the earlier work on adopted-away children of schizophrenic mothers using modern diagnostic techniques, which are more standardized and therefore more reliable than those used in the past. He found that schizophrenia developed in 8 per cent of the adopted-away children of schizophrenic mothers compared with 1 per cent of adopted-away children of psychiatrically healthy mothers. On the other hand, when Tienari examined the proportion of spectrum disorders in the adopted-away children, his results did not support a genetic aetiology.

Puzzle

The solution to the puzzle of schizophrenia has been approached from a multiplicity of directions. Following the recognition of the antipsychotic activity of chlorpromazine in 1954, research focused on the dopaminergic neurons of the brain. Dopamine blockade has now been established as the central mode of action of antipsychotic drugs. Although this appeared to be a promising approach to establishing the biochemical abnormality in schizophrenia, the 'dopamine hypothesis' has been extensively investigated for several decades without the hoped-for breakthrough. It seems likely that dopaminergic abnormalities will turn out to be secondary to some more fundamental disturbance in brain function or structure.

Studies of the neuroanatomy of the brain had been out of fashion with schizophrenia researchers until the recent development of sophisticated imaging techniques. A long catalogue of brain-image abnormalities has now been built up, including reduced volume of the brain in frontal regions, the temporal lobe, the hippocampus and the cerebellum, and abnormalities of the corpus callosum³. None of these is specific to schizophrenia; only a small proportion of patients exhibits any gross abnormality and many patients have brain images that are indistinguishable from those of normal controls⁴. Nevertheless, in the small samples of brains which become available for post mortem examination, some of the abnormalities revealed by imaging have been supported — one study, for example, reporting a reduced volume of the hippocampus in the brains

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

In London's Trafalgar Square — a myth or a diagnostic entity?

of 18 schizophrenic patients compared with 21 control subjects⁵.

Studies of patients with differing durations of schizophrenia suggest that any structural abnormalities of the brain are not progressive. This observation has generated hypotheses that either damage is done around or before the time of birth, or that there is a disturbance in the later development of the brain possibly as late as adolescence⁶ as the disease is extremely rare before the age of 15. Hypotheses concerning structural or developmental abnormalities of the brain have to accommodate the fact that most schizophrenic illnesses run an episodic course, and that a proportion of sufferers lead normal lives between episodes.

One curious observation, initially recorded decades ago and replicated many times since, is that more people who develop schizophrenia are born in the late winter or spring, the excess being about 7 per cent⁷. This has led to the search for a virus, and the elaboration of theories about viruses impeding the migration of neurons through the fetal brain, or being incorporated in genes. Because virus infections wax and wane over time, it is not surprising that attempts have been made to associate those fluctuations with temporal variations in the appearance of schizophrenia, even to the extent of a link between influenza epidemics and an excess of births of people who later develop schizophrenia⁸.

A handful of theorists even postulates that schizophrenia is a recent disease, appearing in the eighteenth century in response to a new infectious agent, and

reaching epidemic proportions in the nineteenth century, thus accounting for the vast building programme of asylums. Such theorists have been greatly encouraged by recent reports of a decline in the incidence of schizophrenia from the few centres that maintain case registers⁹. Controversy rages over whether this indicates a real reduction in prevalence of the disease or whether it is merely a consequence of changing definitions and patterns of service provision. However, other registers have recorded a steady incidence over the years. Furthermore, an extensive epidemiological study by the World Health Organisation showed that if a narrow definition of schizophrenia was applied the incidence was remarkably similar across cultures and climates¹⁰.

By contrast, a series of WHO studies has revealed that the outcome of schizophrenia is decidedly better in developing countries, despite scant follow-up treatment, than in the West¹¹. This appears partly to be due to greater tolerance and acceptance of patients by families in less urbanized and industrialized societies¹². Although playing no part in aetiology, family attitudes have been shown to influence the course of schizophrenia once it has appeared. This finding has stimulated the development of methods of helping families to cope better with the illness, by improving their understanding of its nature and their ability to tackle the daily problems it generates.

Discussions about the nature of schizophrenia commonly throw up the Chinese parable of the blind man

attempting to describe an elephant by feel. In the case of schizophrenia, a more appropriate animal would be the chimaera, the composite mythological beast. Indeed there is a growing body of opinion that schizophrenia is an umbrella term covering several diseases of differing aetiology. Because this is a distinct possibility, it seems wise to retain a wide boundary, within which the phenomena of illness are defined as carefully as possible. There is also a small but vociferous group that continues to claim that schizophrenia is a myth: a label used by psychiatrists acting as society's thought-police to disempower its deviant and politically embarrassing members.

At the centre of this labyrinth of hypotheses lurks a particularly vicious form of catch 22. The boundaries of schizophrenia cannot be defined in the absence of a pathological test, but it is very difficult to identify such a test when the condition is so ill-defined. There is no doubt, however, that whether one applies a broad or a narrow definition, schizophrenia constitutes one of the world's most pressing public health problems. The run-down of the Victorian asylums in Europe and the United States has made schizophrenic patients much more familiar to us all. In the United States, in particular, this occurred with great speed, most patients being discharged to private landlords with no expertise in the care of the mentally ill. It was inevitable that many psychotic patients would end up living on the streets, a situation that has become a public scandal in the larger US cities. Well-planned and well-resourced programmes of psychiatric hospital closure in Britain have shown that such scandals are avoidable. Nevertheless, there are still enough casualties of the system for anyone to form an opinion as to whether schizophrenia is a myth or a diagnostic entity by holding a conversation with their local bag lady or man. □

Julian Leff is Director of the MRC Social and Community Psychiatry Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK.

1. Sherrington, R. *et al.* *Nature* **336**, 164–167 (1988).
2. Aschauer, H., Aschauer-Treiber, G., Cloninger, C. R., Garver, D. L. & Isenberg, K. E. *Schiz. Res.* **2**, 39 (1989).
3. Shelton, R. & Weinberger, D. R. in *The Neurology of Schizophrenia* (eds Nasrallah H. A. & Weinberger D. R.) (Elsevier, Amsterdam, 1986).
4. Seidman, L. J. *Psychol. Bull.* **94**, 195–238 (1983).
5. Bogerts, B. *et al.* *Schiz. Res.* **3**, 295–301 (1990).
6. Benes, F. M. *Schiz. Bull.* **15**, 585–593 (1989).
7. Boyd, J. H., Pulver, A. E. & Stewart, W. *Schiz. Bull.* **12**, 173–186, (1986).
8. O'Callaghan, E., Sham, P., Takei, N., Glover, G. & Murray, R. M. *Lancet* **337**, 1248–1250 (1991).
9. Der, G., Gupta, S. & Murray, R. M. *Lancet* **335**, 513–516 (1990).
10. Sartorius, N. *et al.* *Psychol. Med.* **16**, 909–928 (1986).
11. World Health Organization *Schizophrenia: An International Follow-up Study* (Wiley, Chichester, 1979).
12. Leff, J. P. *et al.* *Brit. J. Psychiat.* **151**, 156–173 (1987).

The working of vitamin K

A neat analogue of vitamin K (essential for blood clotting) points to the likely mechanism for the vitamin's activity — and to the chemical complexities with which evolution has lumbered itself.

WIDESPREAD addiction to the swallowing of vitamin pills appears not to be seriously impeded by the general ignorance of how these relatively simple chemicals can have such profound effects on human and, more generally, animal health. The usual assumption is that each vitamin functions as a cofactor in a chemical reaction catalysed by a naturally occurring enzyme, but that does no more than make the conundrum seem more recondite. Luckily, for at least vitamin K, which is essential for the normal clotting of mammalian blood, there is now a great deal more to say. Since 1974, it seems to have been clear that the function of vitamin K in blood-clotting is to convert the dicarboxylic amino acid glutamic acid (Glu) into γ -carboxyglutamic acid (abbreviated as 'Gla'), which differs from Glu by the substitution of a second carboxyl group at the carbon atom furthest away from the amino group. By now, there are physiological experiments to suggest why this is important. The view that has emerged is that the Glu amino acids in the blood-clotting proteins are converted, under the influence of vitamin K, into Gla residues, which are efficient at binding to calcium ions. Then, the argument goes, the modified clotting proteins and their calcium bind in unspecific ways to phospholipid molecules on the surfaces of platelets in the blood and, hey presto!, there is a blood clot. Several proteins in the clotting cascade are known to function properly only if they are first activated by vitamin K and bound to calcium ions. The same precondition may be required in other physiological processes, bone calcification for example.

The relevance of the conversion of Glu to Gla is easily imagined: the two carboxyl groups attached to a single carbon atom in Gla will neatly form a cage in which a calcium ion can be bound. Just why the proteins then so readily cement together with platelets to form a clot is necessarily a much more complicated matter. But the mechanism by which vitamin K turns Glu into Gla is far from self-evident.

But Paul Dowd, Seung Wook Ham and Steven J. Geib from the University of Pittsburg now have some novel data on which to base a new view of the mechanism of this unusual reaction (*Am. chem. Soc.* **1991**, 7734; 25 September 1991). What they have to say, its own interest apart, is yet another reminder of how the process of evolution seems to have been heavily dependent on the chemical

reactions that happened to be available when they were needed, and has been forced to live with them ever since.

Carboxylation is in itself an odd business, being one of the ways in which an extra carbon atom can be added to a molecule. The essential question is that of why vitamin K should be such an effective handmaiden of the process. Whatever the process, it is clear that the vitamin K must be regenerated (for otherwise those at risk from blood-clots would have to swallow grams of it each day). Free oxygen and free carbon dioxide are also essential for the process.

The architecture of the molecule is built on 2-methyl-1,4-naphthoquinone (a naphthalene molecule, consisting of two benzene rings sharing a common side, in which one of the rings carries a methyl group and is also a quinone, having oxygen atoms substituted for hydrogen atoms at opposite ends). To the unsubstituted position in the quinone ring is attached a long 20-carbon hydrocarbon side chain — a phytyl group will give vitamin K₁, but other variants are as effective. It seems well-established that the active form of vitamin K is that in which the quinone is reduced to a hydroquinone (which means that the double-bonded oxygen atoms become hydroxyl groups). In real life, an enzyme (called naturally a carboxylase) is also required. What Dowd and his colleagues have done is to dispense with the enzyme as well as with vitamin K itself. Their goal is to simulate the real-life reaction with with an analogue that will more conveniently illustrate the mechanism. The reasoning is that carboxylation must at the least require the removal of a hydrogen atom from the position at which the extra carboxyl group is to appear. That calls for a base in the strict sense — a proton acceptor. Then, with luck, the addition of the carboxyl group will entail no more than the attachment of a molecule of CO₂.

The analogue system is ingeniously devised. Dowd and his colleagues started with a naphthoxide (which is a naphthalene molecule carrying a phenol group and two methyl groups in the *ortho* and *para* positions on the same ring. Anions of this acid (in the unstrict sense) react spontaneously with molecular oxygen to form an alkoxide base (the product of the addition and rearrangement of an oxygen molecule) and that proves strong enough a base to extract a hydrogen atom from another molecule and so to turn an aliphatic diester into a ring molecule.

That, of course, is not the same thing as the addition of another carboxyl group, but it is enough to show that a molecule similar to, if simpler than, vitamin K will, under the influence of molecular oxygen alone, form a sufficiently strong base to drive the extraction of protons from the position in a simple amino acid molecule next to a carboxyl group. And the molecules are simple enough for there to be data showing that the scheme should function energetically. It needs only a little imagination to see that this may also happen with vitamin K. The authors cheerfully combine speculation and some hard-won data to argue that this is how vitamin K functions in real life.

So how much wiser are we, then? A little or a lot? That is almost a test-question for a public examination in the management of research. Politically cautious, or correct, candidates will say there will be a substantial pay-off, undoubtedly greater than the initial cost of the research, that may nevertheless not be realised in a lifetime. It is easy to imagine, for example, that minor changes in the chemical structure of the quinone might make vitamin K a more effective product, or that a little fiddling with the phytyl side-chain might change the specificity of the enzymic co-factor.

But that is only a small part of the whole story. It says a great deal for the sophistication of an ancient craft (chemistry) that people should be able to convince others of the aptness of their speculations by creating model systems that merely simulate what may happen in real life. The work is even more dangerous than building a Global Climatic Model. Its practitioners deserve praise.

Meanwhile, there remains the evolutionary point: vitamin K is evidently a molecule which can react with molecular oxygen to form a powerful base. That allows it to trigger the blood-clotting system, for which reason its structure has been conserved. If the slings and arrows of the environmental history of mammals had been different, no doubt there would have been a different handmaiden of the blood-clotting process. It would be interesting, and probably important as well, to know more about these differences and their origins. Probably the place to look for explanations is in the comparison of biochemical mechanisms in mammals still extant. The people who engage in that pursuit may be called palaeobiochemists, but all their material will be taken from the wild.

John Maddox

It may be better to be a wimp

R. McNeill Alexander

AMONG the mammals, the champion endurance athlete is the antelope-like pronghorn from the North American prairies. That is the conclusion of Lindstedt *et al.* (page 748 of this issue¹), who took pronghorn to the limit of their performance by running them uphill on a sloping treadmill.

Lindstedt and colleagues found that pronghorn can use oxygen 3.3 times as fast as predicted for a typical mammal of their 32 kilogramme mass, at a rate corresponding to metabolism at 100 watts per kilogramme. The maximum rate of oxygen consumption sets the maximum speed of running that can be sustained for more than a few hundred metres. The new data show that, on level ground, pronghorn should be able to sustain over 20 metres per second (45 miles per hour), and there is a field observation of pronghorn running 11 kilometres in 10 minutes (18 m s⁻¹). Most horse races are won at only 16–17 m s⁻¹.

Pronghorn seem to achieve their remarkable performance by very simple means. They have large lungs with exceptionally high diffusing capacity. Their blood has a high haemoglobin content and their cardiac output is high. Their muscles make up a larger than usual fraction of body mass and are exceptionally rich in mitochondria. The pronghorn's outstanding running capacity does not depend on any evolutionary novelty, but simply on enhancement of normal mammalian structure and physiology. Why have other mammals not evolved to match it?

Similar questions could be asked about other animal champions. The Australian hopping mouse (*Notomys*) conserves water in its desert habitat by producing more concentrated urine than any other mammal that has been investigated². To do this it needs very long kidney tubules, but why have these not been matched by other mammals? Tortoises have leg muscles that are exceptionally economic of energy³, and the armoured shrew (*Scutisorex*) has a backbone strong enough to support the weight of a man⁴. Would other animals not benefit from more economical muscles and stronger skeletons?

The answer to these questions is surely that exceptional performance is accompanied by inevitable costs. In some cases the nature of the costs seems clear: the economy of tortoise muscles has been gained at the expense of speed, and strong skeletons are cumbersome. It is harder to identify the costs for the pronghorn. We might expect its resting meta-

bolic rate to be rather high for various reasons (for example, its haemoglobin-rich blood may be unusually viscous, so the energy cost of pumping it around the body may be high), but measured metabolic rates are about the same as for other ungulates of similar size⁵.

Whatever the costs may be, there is likely to be an optimum maximum rate of oxygen consumption for any mammal, determined by the balance between costs and benefits. An analogy from engineering can help to make this clear.

An engineer might decide how strong to build a structure by reckoning the costs of construction, and the costs of insurance against failure, for designs of different strengths. (Stronger structures cost more to build but are cheaper to insure.) The optimum design would minimize the total cost. Such an approach has been used in a theory of optimum strength for skeletons⁶ and might be applied in a theory of optimal capacities for oxygen uptake: the increased cost of improved running performance would be weighed against the reduced premium for a notional insurance policy against being caught by a predator.

It would be formidably difficult to assess the premium and the costs in the same currency, but the approach may nevertheless be useful. Tortoises can

COLON CANCER

Suppression with a difference

Henry R. Bourne

THE latest tumour-suppressor gene to be identified, *APC* (for adenomatous polyposis coli), has been reported simultaneously by research groups writing in *Science*^{1,2} and *Cell*^{3,4}. The gene encodes a big protein that probably forms protein-protein complexes bound together in 'coiled coils'. It is located in region 21 of the long arm of chromosome 5 (5q21), is expressed in many different cell types and is subject to somatic mutations in sporadic colorectal tumours.

"*Déjà lu*, or something new?" asks the reader, remembering the *MCC* (mutated in colorectal cancer) gene reported barely five months earlier⁵. The *APC* gene does remarkably resemble *MCC* and the two are near neighbours at 5q21, but there is a key difference (see Table): individuals who inherit germ-line mutations in *APC* develop hundreds of colonic polyps at an early age; without treatment, all eventually come down

benefit from economical but slow muscles because their armour reduces the risk of damage by predators that they cannot outrun. Some moas (extinct ostrich-like birds from New Zealand) had extraordinarily thick leg bones, possibly because the absence of predators shifted the balance of advantage between the agility allowed by light bones and the strength of heavy ones⁷. Similarly, the exceptional running performance of pronghorns might be explained by the threat from the wolf, a pursuit predator, rather than from the cats and other ambush predators that predominate in many other habitats⁸, although for other mammals the balance of advantage probably favours less extreme ability. For these other mammals, it may be better to be an inferior performer — a thought that less athletic readers may find consoling. □

R. McNeill Alexander is in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.

1. Lindstedt, S. L., Hokanson, J. F., Wells, D. J., Swain, S. D., Hoppeler, H. & Navarro, V. *Nature* **353**, 748–750 (1991).
2. MacMillen, R. E. & Lee, A. K. *Comp. Biochem. Physiol.* **28**, 493–514 (1969).
3. Woledge, R. C., Curtin, N. A. & Homsher, E. *Energetic Aspects of Muscle Contraction* (Academic, London, 1985).
4. Allen, J. A. *Bull. Am. Mus. nat. Hist.* **37**, 769–784 (1917).
5. McNab, B. K. *Ecol. Monogr.* **56**, 1–19 (1986).
6. Alexander, R. McN. *J. theor. Biol.* **109**, 621–636 (1984).
7. Alexander, R. McN. *J. Zool., Lond.* **201**, 363–376 (1983).
8. Van Valkenburgh, B. *Palaeobiology* **11**, 406–428 (1985).

with colorectal cancer. In contrast, no germ-line mutations in *MCC* have been found. The difference carries implications for the basic understanding of tumour suppression and for families affected by the inherited disorder, familial adenomatous polyposis (FAP).

The discoveries of *MCC* and *APC* emerged from a mix of collaboration and competition among three laboratories, led by Bert Vogelstein in Baltimore, Ray White in Salt Lake City, and Yusuke Nakamura in Tokyo. Mapping of the FAP defect to 5q21, plus deletions of 5q21 DNA in colorectal tumours, led first to the *MCC* gene. In March, a collaborative report⁵ from all three groups described *MCC* deletions and point mutations in sporadic colorectal tumours.

The proposal⁵ of *MCC* as a candidate for the inherited defect in FAP proved a dud, however — no germ-line *MCC* mutations have yet been found in FAP

patients², and small deletions of neighbouring DNA found in several FAP patients did not include *MCC*⁴. Instead, further searching produced compelling evidence that FAP results from inheritance of mutations in a nearby gene: this gene, *APC*, suffers germ-line mutations in FAP patients, and virtually all the mutations inactivate the gene (see Table). In one family, a stop codon that truncates the *APC* coding region cosegregated precisely with the FAP phenotype in 21 family members, including 12 patients with FAP².

How does inheritance of an inactivating *APC* mutation cause cells of the colonic mucosa to proliferate abnormally and form polyps (benign adenomas)? Each polyp in an FAP patient probably requires, in addition, a second genetic event, a somatic mutation in a single mucosal cell. This explains why only a few hundred of the millions of cells in the colonic mucosa of an FAP patient give rise to polyps, and why even these do not appear until the second and third decades of life. This genetic explanation of polyp formation parallels our understanding of neoplastic progression in general, and of colorectal tumours in particular. Genetic analysis suggests that emergence of sporadic (non-familial) colorectal carcinomas from polyps correlates with accumulation of mutations in several genes, including *ras* and three tumour-suppressor genes^{5,6} (encoding p53, a putative cell-surface adhesion protein and *MCC*). Now the list also includes *APC*, which is mutated in some sporadic colorectal tumours² but unaffected in other cells of the same patients (see Table).

The postulated 'second hit' mutation in polyps of FAP patients does not appear to strike the normal *APC* allele: polyps of FAP patients do not have deletions at 5q21 in the 'normal' chromosome. This situation differs from that in retinoblastoma, in which patients who inherit an inactivating mutation in *Rb*, another tumour-suppressor gene, develop malignant retinoblastomas only when a second mutation inactivates the remaining normal *Rb* allele. Thus *Rb* mutations behave as 'recessive' in individual retinal cells, whereas inherited *APC* mutations act in a dominant fashion in individual mucosal cells.

To explain why a heterozygote exhibits a mutant phenotype, geneticists invoke either of two standard explanations. In the first, 'haplo-insufficiency', one normal copy of the gene simply does not suffice to support normal function. In this view, the normal complement of *APC* protein (produced from two normal *APC* alleles) enables a mucosal cell to resist neoplastic transformation (generation of a polyp) triggered by the postulated 'second hit'; half that amount

Genes at chromosome 5q21 mutated in FAP and colon cancer

	<i>MCC</i>	<i>APC</i>
Protein	829 amino acids, with coiled coils scattered throughout its length ⁵	2,843 amino acids, with coiled coils in N-terminal 25 per cent ^{1,4}
Somatic mutations (in tumours)	6 point mutations in 90 tumours (about 35 per cent of coding region examined) ² Includes 2 truncating mutations*, 4 amino-acid substitutions	3 point mutations in 158 tumours (about 5.4 per cent of coding region examined) ²
Germ-line mutations in patients with FAP	None ^{2,3} (about 35 per cent of coding region examined in 90 FAP kindreds ²)	Inherited with disease in families ^{2,3} . Of 9 mutations [†] in 164 unrelated patients, 8 truncate the protein* and 1 substitutes an amino acid ^{2,3}

* A truncated protein results from substitution of a stop codon, an insertional or deletional frameshift, or alteration of splice junction.

[†] Although the number of mutations seems quite small, this may be misleading: so far, only a small part of the *APC* coding region (less than 10 per cent, see ref. 2) has been surveyed for mutations.

of protein, however, is not enough. Any satisfactory explanation of the quasi-stoichiometric tumour-suppressing ability of *APC* will need an understanding of what *APC* actually does in cells.

In the meantime, it is easy to imagine that *APC* acts as a brake on cell division by binding to and controlling the activity of a cell component whose concentration regulates proliferation (explaining how a 50 per cent decrease in *APC* could render a cell especially susceptible to oncogenic mutations). Different amounts or activities of this second component in different cell types could account for a curious feature of FAP: despite the expression of *APC* in many tissues, inheritance of defective *APC* produces a phenotype that is usually limited to colonic mucosa (although some patients suffer from tumours at other sites as well). Human genetic diseases provide many examples of haplo-insufficiency — for example, familial hypercholesterolaemia and type-1 pseudohypoparathyroidism, caused by a 50 per cent deficiency of the low-density-lipoprotein receptor and of the α subunit of G_s protein, respectively. In both disorders, inactivation of one allele of a widely expressed gene causes functional defects limited to a small number of tissues.

A second explanation for mutant behaviour of the heterozygote is that the mutant protein exerts a 'dominant negative' effect⁷. Both the Baltimore-Tokyo^{1,2} and the Utah^{3,4} groups raise this possibility, based on predicted coiled-coil structure in the amino-terminal quarter of the *APC* protein. (For a brief account of coiled-coil structures, and how they can be predicted from amino-acid sequence, see my recent News and Views article⁸.) The idea is simple: one domain of *APC* partici-

pates in coiled-coil complexes with other proteins, while the remaining three quarters of the protein performs other functions, which depend upon formation of the complexes. A mutation that removes or inactivates the latter domains but preserves the ability to form complexes may produce a protein that can bind to and inactivate normal components of the complex. Thus a small number of mutant protein *APC* molecules (50 per cent or less) could produce a disproportionately large dominant negative defect in function — as already demonstrated for myosin mutations in worms⁹ and man^{10,11}. In keeping with this idea, most of the reported mutant *APC* genes would produce a protein of at least 200 amino acids, including predicted coiled-coil regions.

A few months ago⁸, I applied this argument to the wrong protein, *MCC* (*déjà lu*, indeed). The idea is probably good enough for the right protein as well. As the discoverers of the two genes suggest¹⁻⁴, it is even possible that *MCC* and *APC* normally coil with each other in cells of the colonic mucosa — thus explaining, in one grand wave of the hand, why mutations in either protein can occur in colorectal tumours, and why 5q21 deletions are found so frequently in the same tumours. ►

1. Kinzler, K. W. et al. *Science* **253**, 661–665 (1991).
2. Nishisho, I. et al. *Science* **253**, 665–669 (1991).
3. Groden, J. et al. *Cell* **66**, 589–600 (1991).
4. Joslyn, G. et al. *Cell* **66**, 601–613 (1991).
5. Kinzler, K. W. et al. *Science* **251**, 1366–1370 (1991).
6. Fearon, E. R. & Vogelstein, B. *Cell* **61**, 759–767 (1990).
7. Herskowitz, I. *Nature* **329**, 219–222 (1987).
8. Bourne, H. R. *Nature* **351**, 188–190 (1991).
9. Bejsovec, A. & Anderson, P. *Genes Dev.* **2**, 1307–1317 (1988).
10. Tanigawa, G. et al. *Cell* **62**, 991–998 (1990).
11. Geisterfer, L. A. et al. *Cell* **62**, 999–1006 (1990).

Finally, a less grand but more satisfying prediction: discovery of *APC* will benefit real patients, those individuals (about 1 person in 8,000) afflicted with familial adenomatous polyposis. Specific identification of an *APC* mutation in a family will allow early diagnosis and treatment of related individuals who inherit the defective gene. Relatives who do not carry the abnormal gene will

also benefit by avoiding both the anxiety caused by knowing their family is cancer-prone and the discomfort of repeated examinations. □

Henry R. Bourne is in the Departments of Pharmacology and Medicine and the Cardiovascular Research Institute, University of California, San Francisco, California 94143, USA.

PHOTOELECTROCHEMISTRY

Bettering nature's solar cells

Thomas E. Mallouk

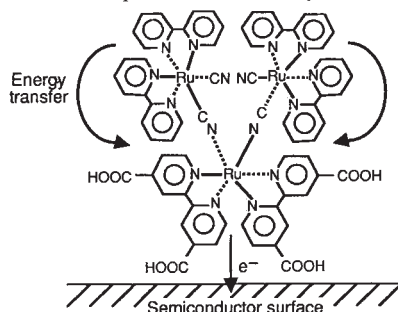
A NEW solar cell, described by Brian O'Regan and Michael Grätzel on page 737 of this issue¹, surpasses the performance of the natural solar cells used in photosynthesis. The authors' trick, borrowed from nature, is to use a light-harvesting 'antenna' complex to direct excitation to a semiconductor-based photosynthetic system.

In the wake of the discovery, twenty years ago², that water could be decomposed at the surface of illuminated oxide semiconductor electrodes, most electrochemists were optimistic that the development of inexpensive liquid-junction solar cells based on this process was just around the corner. In principle, highly efficient cells could be made from semiconductors such as Si, GaAs, InP, CdTe and WSe₂, the bandgaps of which, 1.3 ± 0.3 electronvolts (eV), overlap well with the solar spectrum. Although liquid-junction devices made from pure single crystals of these materials are indeed efficient, in fact slightly better than analogous solid-state Schottky junction cells³, inexpensive polycrystalline photoelectrodes rarely give solar power conversion efficiencies in excess of 5 per cent. With polycrystalline materials, grain boundaries trap minority charge carriers and seriously undermine the performance of the device.

A way to avoid this problem is to adsorb onto the surface of the polycrystalline electrode a light-absorbing photosensitizer molecule, which can inject only majority carriers. Such a device is the one described by O'Regan and Grätzel. Their cell consists of relatively inexpensive materials — high-surface-area, polycrystalline TiO₂, with a monolayer coating of the complex shown in the figure — and has a solar-to-electric conversion efficiency of 7–12 per cent, depending on conditions. In terms of cost and efficiency, their device is comparable to the best amorphous-silicon p-i-n solid-state cells produced today.

The idea of sensitizing wide-bandgap semiconductors to make them responsive to visible light is also about twenty years old. Early experiments by Gerischer and

co-workers used solution-phase dyes such as rhodamine B to sensitize oxide semiconductors⁴. The problem in this case was that the excited-state lifetime of a strongly absorbing dye is typically 10^{-8} – 10^{-7} sec, so that excited, dissolved molecules would lose their energy long before diffusing to the electrode surface. Surface-adsorption of the sensitizer appeared only to make matters worse⁵: even if the quantum efficiency for charge



Energy and electron flows within the sensitizer molecule designed by Amadelli *et al.*⁷ and used by O'Regan and Grätzel¹.

injection from the molecule is close to unity (as it is in some special cases⁶), a monolayer of even the most highly coloured dye absorbs less than 1 per cent of the incident light. Remarkably, the TiO₂ electrodes used by O'Regan and Grätzel have a roughness factor of 200, largely compensating for this effect, and yet they are sufficiently conducting that photoinjected electrons are collected from the entire porous film. The light-harvesting efficiency is further enhanced by the use of the triruthenium complex (see figure), which has two 'antenna' ruthenium bis(bipyridyl) moieties funneling excitation to the ruthenium bis(dicarboxybipyridyl) sensitizer⁷. The antennae and sensitizer absorb blue and green light, respectively, improving the overall spectral response.

What are the prospects for making more-efficient low-cost devices, using this or other designs? The quantum yield for charge injection by the triruthenium dye is practically unity, and the excess free energy of the excited-state sensitizer is 1.63 eV per molecule⁷. Clearly the

quantum yield cannot be improved; however, the observed photovoltage is only about 0.58 V at the point of maximum power. The missing 1.05 V is thermalized in driving the reaction at the optimal rate. About 0.4 V is given up in transferring an electron from the sensitizer to TiO₂. Another 0.4 V is lost in moving the electron from the outer surface of the TiO₂ electrode to the counterelectrode, and in reducing triiodide to iodide. Finally, about 0.2 V is used to transfer an electron from iodide to the oxidized sensitizer⁸. Of these losses the first could be reduced significantly, by fabricating cells from other semiconductors with more negative conduction band edge potentials. Such semiconductors (for example, SrTiO₃, MgTiO₃ and CaTiO₃) can in fact be sensitized by ruthenium(dicarboxybipyridyl) complexes and other dyes⁹, but whether conducting, high-surface-area electrodes can be made from them remains to be seen.

It is interesting to compare this device with the natural photosynthetic apparatus; indeed, there are several parallels. First, green plants also use a light-harvesting antenna system, to direct excitation energy to chlorophyll molecules in photosystems I and II. Second, much of the energy absorbed is sacrificed to keep the quantum yield for charge transfer near unity. In green-plant photosynthesis, two 1.80-eV photons are used, per electron, to drive the 1.24-eV per electron reduction of CO₂ to sugars. Finally, the power conversion efficiencies are comparable. The overall efficiency for CO₂ reduction, in terms of free energy conversion, has been estimated to be 9 per cent *in vivo*¹⁰; but because plants must consume energy to support metabolic processes, the actual free energy available from biomass is less than half this. In this sense the solar cell of O'Regan and Grätzel, developed after two decades of intensive work on photosensitization, is better than the product of a billion years of evolution.

In defence of biological photosynthesis, it should be pointed out that natural systems offer some other clear advantages as energy-conversion devices. First, they are self-repairing, reproduce themselves and require little in the way of wiring or support structures. Second, through rather exquisite biochemical pathways, they produce fuels from abundant feedstocks such as water and CO₂, rather than electricity. In some cases very energy-rich fuels are made directly.

Interesting examples are unicellular algae which, starved of CO₂, are stimulated to produce the enzyme hydrogenase and release their reducing equivalents as molecular hydrogen¹¹. Certain plants, such as *Copaifera multijuga*, a tropical tree, can be tapped directly for sesquiterpene oils which re-

semble diesel fuel¹². Unfortunately, the named species does not grow in cooler climates. Calvin¹² has suggested that the genes for biosynthesis of sesquiterpenes from *Copaifera* be transferred to species such as *Euphorbia lathyris*, a fuel-producing plant which can be grown in arid regions of the United States. Likewise, one might imagine that eukaryotic algae could be modified to produce more hydrogenase, and to bind CO₂ less effectively at normal partial pressures.

In the light of the rapid progress made within the past five years in eukaryotic gene expression, and indeed in the engineering of entire metabolic pathways in prokaryotic organisms¹³, these semi-natural alternatives to semiconductor-based solar cells could appear in the not so distant future. □

Thomas E. Mallouk is in the Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712, USA.

- O'Regan, B. & Grätzel, M. *Nature* **353**, 737–740 (1991).
- Fujishima, A. & Honda, K. *Nature* **238**, 37–38 (1972).
- Abrahams, I. L. et al. *New J. Chem.* **11**, 157 (1987).
- Gerischer, H. *Photochem. Photobiol.* **16**, 243 (1972).
- Spitler, M. T. & Calvin, M. *J. chem. Phys.* **66**, 4294–4305 (1977).
- Spitler, M. T. & Parkinson, B. A. *Langmuir* **2**, 549–553 (1986).
- Amadelli, R., Argazzi, R., Bignozzi, C. A. & Scandola, F. *J. Am. chem. Soc.* **112**, 7099–7103 (1990).

- Roffia, S. et al. *J. electroanal. Chem.* **302**, 157–171 (1991).
- Hashimoto, K., Hiramoto, M. & Sakata, T. *Chem. Phys. Lett.* **148**, 215–220 (1988).
- Bolton, J. R., Strickler, S. J. & Connolly, J. S. *Nature* **315**, 495–500 (1985).
- Greenbaum, E. in *Photochemical Energy Conversion* (eds Norris, J. R. Jr & Meisel, D.) 184–195 (Elsevier, New York, 1989).
- Calvin, M. *J. chem. Ed.* **64**, 335–336 (1987).
- Bailey, J. E. *Science* **252**, 1668–1675 (1991).

PHYSIOLOGY

Cold facts and naked truth

Andrew R. Cossins

THAT curious mammal, the naked mole rat, subject of a recent biography¹, is even odder than it had seemed. Writing in the *Journal of Thermal Biology*², Buffenstein and Yahav report that the animal is unable to regulate its body temperature and instead displays all the characteristics of a poikilotherm.

Not so long ago there were only two types of animal, warm-blooded (birds and mammals) and cold-blooded (the rest). This simple classification became confused by the discovery of warm-blooded insects and reptiles, and of periodic hypothermia in small mammals³. Nevertheless, until now mammals and birds could still be characterized by their high resting metabolic rates, an effective surface insulation and high body temperatures in the cold.

The naked mole rat, *Heterocephalus glaber*, is found only in the tropical, semi-arid zones of north-east Africa, where it lives in a complex of poorly ventilated, subterranean burrows. Food is sparse and foraging through the burrow is an activity that dominates the energy budget. On the other hand, the burrows are thermally very stable at 28–32 °C but with a high humidity and the potential for low levels of oxygen and high levels of carbon dioxide⁴. The animals have a complex social structure. They live in subterranean colonies of 60–100 individuals, with a strict division of labour extending even to the provision of a single breeding female⁵. They often rest together in large huddles up to four animals deep.

The most notable features from a thermoregulatory point of view are a complete absence of fur and a poorly

developed subcutaneous fat layer. The skin is loosely folded over the surface and possesses only a few whiskers, mainly around the snout. It lacks sweat glands but is nevertheless highly permeable to water. The absence of an effective insulation layer, together with the animals' small size and low basal metabolic rate, were thought to restrict their thermoregulatory capacity and they were commonly regarded as poorly regulating homeotherms⁶.

However, it is now clear that they also lack another fundamental attribute of the small homeothermic mammal, an effective mechanism for increasing the production of metabolic heat in the cold. One obvious manifestation of this property is an ability to elevate body temperature above a cool ambient temperature, but Buffenstein and Yahav show that the body temperature of the naked mole rat is only slightly (0.5 °C) above ambient over the range 12–37 °C. Another test of thermogenic potential is an animal's increased rate of oxygen consumption (and heat production) when transferred to the cold via the rapid activation of brown adipose tissue. Buffenstein and Yahav show that oxygen consumption is in fact greatly reduced in the cold.

This lack of thermogenic potential separates naked mole rats even from hibernating mammals of similar size, their production of metabolic heat being more than compensated for by high rates of evaporative and conductive cooling².

The lesson here, however, is that the thermal conditions under which our present thermoregulatory taxonomy has been developed are irrelevant to these animals, because they never (or hardly ever) experience ambient temperatures below 28 °C in their burrows⁴. So although they are technically poikilotherms⁷, their body temperature never falls below 28 °C.

The naked mole rat therefore provides an exception to the rule that resting mammals have an effective system for endothermic thermoregulation. However, it is not simply a degenerate mammal in this respect, but has probably sacrificed a redundant physiology for some other adaptive benefit. It may be that reduced metabolic rates lower the possibility of overheating in an enclosed and humid burrow. Or perhaps the energetic savings on abandoning endothermy are crucial to balancing the energy budget in a food-scarce habitat.

Either way these animals offer new opportunities for the study of hypothermia in mammals because, unlike every other mammal, body temperature can be set at any desired level, at least in the short term, as it can in poikilotherms. This convenience may be also appreciated in studies of cold preservation of transplantation organs, where tissue may suffer from hypothermia *in vivo*. Experimental biology is littered with

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Naked mole rat: new insights into hypothermia

examples of unusual creatures providing insight into fundamental biological processes. So, for the naked mole rat, opportunity knocks. □

Andrew R. Cossins is in the Department of Environmental and Evolutionary Biology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

- Sherman, P. W., Jarvis, J. & Alexander, R. D. (eds) *The Biology of the Naked Mole-Rat* (Princeton University Press, 1991).
- Buffenstein, R. & Yahav, S. *J. Thermal Biol.* **16**, 227–232 (1991).
- Cossins, A. R. & Bowler, K. *Temperature Biology of Animals* (Chapman & Hall, London, 1987).
- Bennet, N. B. et al. *S. Afr. J. Zool.* **3**, 189–195 (1988).
- Jarvis, J. U. M. *Science* **212**, 571–573 (1981).
- Withers, P. C. & Jarvis, J. U. M. *Comp. Biochem. Physiol.* **66A**, 215–219 (1980).
- IUPS Thermal Commission *Pflügers Arch. ges. Physiol.* **410**, 567–587 (1987).

Country life

A MAGNIFICENT west-facing victorian country residence with two spacious outbuildings in charming mature woodland, conveniently adjoining a golf course on the English South Downs, was demolished recently to make way for a new road. With 60 rooms, more than a kilometre of corridors and no fewer than 225 doors, the entire property was subterranean and owned by a family of badgers (*Meles meles*). T. J. Roper *et al.* took advantage of the need to check that the badgers had vacated the premises before the road-building and excavated the sett complex, which consisted of one main sett and two auxiliaries. Such opportunities, though, are rare, and there is little idea of the typical size of a badger sett or even an explanation for the do-it-yourself compulsion that drives the inheritors of these already huge domains to extend and improve them further. Full particulars are on view in *J. Zool., Lond.* **225**, 115–124 (1991).

Feel the length

HOW do the filaments in the sarcomere of a striated muscle know when to stop growing? The answer seems to be that the muscle contains measuring rods in the form of extended proteins of enormous length, titin regulating the thick and nebulin in the thin filaments. M. Kruger *et al.* (*J. Cell Biol.* **115**, 97–107; 1991) now come up with the best evidence for this suggestion, using a set of monoclonal antibodies to mark the positions of nebulin epitopes along the sarcomere. The epitopes all remain at a fixed distance from the Z-line (in which the thin filaments are rooted) even when the muscle contracts. Nebulin *in situ* is at least 0.9 μm long; it is clearly an archetype for a molecular ruler.

Neutrino nudge

THE effects of the Z^0 particle, the quarry of CERN's LEP collider, have revealed themselves in a nuclear physics experiment at Rutherford–Appleton Laboratory (B. Bodmann *et al.* *Phys. Lett. B* **267**, 321–324; 1991). The first evidence of the Z^0 , long before it was actually identified, was the observation in 1972 that neutrinos can bounce off electrons even though they have no electric charge. Until then, neutrinos had been seen only in collisions in which they were consumed (or conversely, created), for example in the neutrino-induced reverse β -decay in which chlorine is converted into argon. In the new experiment, a low-energy neutrino passes through a carbon nucleus, but on its way gently nudged the nucleus into an excited state through the 'weak neutral current' mediated by the Z^0 . The authors note with satisfaction that the standard model calculations for the interaction's strength work as well at low energy as they do at high-energy.

Antihydrogen in a new light

Richard Hughes

THE concept of antimatter, introduced by Dirac in 1928, has become so familiar that the use of antiprotons and positrons in high-energy physics experiments is regarded as quite routine. Atomic antimatter, however, has yet to be studied experimentally, largely because of the difficulties involved in producing and manipulating it. But two experimental groups have now reported results^{1,2} that could be applied to the formation of the simplest anti-atom, antihydrogen. Because the spectroscopic techniques that have been developed for hydrogen could be used, this bound state of a positron and an antiproton is the most suitable anti-atom for studying the matter–antimatter and other fundamental symmetries of nature.

The formation of antihydrogen by recombination of a positron with an antiproton has an exact analogue in normal matter: the interaction of an electron and a proton to form hydrogen. This process, which has been studied since the earliest years of quantum mechanics³, requires the participation of a third particle for energy and momentum to be conserved. The simplest reaction is radiative recombination, in which a photon is emitted in conjunction with the formation of the atom. But because photon emission from an atom is a relatively slow process, the probability of this reaction is very small unless the velocities of the participating particles are closely matched. For instance, at CERN's Low Energy Antiproton Ring (LEAR), where 5-MeV antiprotons could be combined with a co-moving, intense positron beam, the production rate would be only a few antihydrogen atoms per second.

In 1983, it was suggested that the reaction rate could be enhanced by a factor of about 100 if the interaction region was illuminated with laser photons with energies equal to those of the photons emitted in the radiative recombination⁴. It is this laser-induced or stimulated recombination that has now been demonstrated with electrons and protons. Schramm *et al.*¹ have observed laser-induced recombination of 21-MeV protons in a storage ring with co-moving 11.4-keV electrons, to form hydrogen atoms in the state with principal quantum number $n = 2$; and Yousif *et al.*² have demonstrated stimulated recombination of protons from a 400-keV Van der Graaf accelerator with a merged beam of electrons, producing hydrogen in the $n = 11$ and 12 levels.

When antihydrogen is eventually produced, the primary focus of experiments

with it is likely to be on precision tests of fundamental symmetries. The experimental techniques that have been developed for the optical spectroscopy of hydrogen could be adapted to this purpose, so remarkably high precisions can be anticipated. For instance, the energy-difference between the ground state and the metastable $2s$ level in hydrogen has been measured with a precision of 5 parts in 10^{10} . Some idea of the precision that is potentially attainable in measurements of this interval comes from noting that the approximately $\frac{1}{8}$ -second lifetime of the $2s$ level translates into an intrinsic quantum-mechanical uncertainty of about 1 part in 10^{15} .

Further improvements in the precision of the hydrogen measurements towards this level will come with developments in optical frequency standards, laser stability and sources of cold atoms. The precision with which the symmetry between matter and antimatter, known as CPT (charge–parity–time), is tested could be improved by several orders of magnitude if such measurements could also be carried out on antihydrogen. This symmetry, which is a consequence of quantum mechanics and relativity, requires that the mass of a particle and its antiparticle be equal, and that their charges be equal and opposite. At present, the masses (really the charge-to-mass ratios) of the proton and antiproton have been shown to be equal to 4 parts in 10^8 , from measurements of their cyclotron frequencies in a magnetic field⁵.

Another fundamental principle of physics that could be studied with antihydrogen is whether antiparticles obey the equivalence principle. This principle, which is the foundation of general relativity, requires that all particles in the same gravitational field fall with equal gravitational acceleration. Although there have been many very precise tests of this principle with matter, there has yet to be an experiment with antiparticles such as the positron or antiproton. A measurement of the gravitational acceleration of antihydrogen would offer the advantage of dealing with an electrically neutral particle, so that the influence of stray electric fields would be reduced. Another way to test this principle would be to look at the gravitational shift of precisely measured transition frequencies in hydrogen and antihydrogen. The equivalence principle requires that the shifts be equal in the two systems, and so a test of equivalence for the positron could be deduced from the results.

The recently demonstrated laser-

induced recombination technique is not the only method for the production of antihydrogen atoms. The successful trapping, storage and cooling of antiprotons to cryogenic temperatures at LEAR offers the potential of producing slowly moving antihydrogen atoms that would be well-suited to high-precision spectroscopic measurements. Two schemes for making antihydrogen have been proposed that capitalize on these developments. In the first of these, antihydrogen would be formed in the reaction between slow-moving antiprotons and cold positronium⁶ (an 'atom' made of an electron and a positron bound together); in the second, cold antiprotons and positrons would be combined in an ion trap⁷. In both cases, the antihydrogen would have to be controlled for subsequent experimental study. Perhaps the ultimate goal would be to trap antihydrogen atoms in the type of magnetic-gradient trap that has been developed for atomic hydrogen. Because of the feeble hold of such traps, the antihydrogen would have to be laser-cooled to exceedingly low

temperatures.

Atomic physics with antimatter had remained a pipe dream until the recent progress in trapping of charged and neutral particles, laser physics and, now, in the production mechanisms. Because of its unique ability to produce a cooled beam of low-energy antiprotons, LEAR remains the most suitable facility for antihydrogen physics. A feasibility study of the fundamental physics that can be studied with antihydrogen is now being pursued at CERN as one component of the LEAR programme beyond 1993. □

Richard Hughes is in the Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

- Schramm, U. et al. *Phys. Rev. Lett.* **67**, 22–25 (1991).
- Yousif, F. B. et al. *Phys. Rev. Lett.* **67**, 26–29 (1991).
- Stueckelberg, E. C. G. & Morse, P. M. *Phys. Rev.* **35**, 116–117 (1930).
- Neumann, R. et al. *Z. Phys.* **A313**, 253–262 (1983).
- Gabrielse, G. et al. *Phys. Rev. Lett.* **65**, 1317–1320 (1990).
- Humbertson, J. W. et al. *J. Phys.* **B20**, L25–L29 (1987).
- Gabrielse, G. et al. *Phys. Lett.* **A129**, 38–42 (1988).

CELL BIOLOGY

Never-in-mitosis in mitosis

Andrew W. Murray

THE cell cycle may be starting to resemble the Soviet Union: the ruling ideology is being pushed aside by a reformist movement from an unexpected quarter. It has been widely believed^{1,2} that the central event in the induction of mitosis is the activation of a protein kinase composed of p34^{cdc2} and mitotic cyclins (p34^{cdc2} is named after the *cdc2* gene that encodes it in the fission yeast, *Schizosaccharomyces pombe*). But experiments from the Osmani laboratory, using the fungus *Aspergillus nidulans*, reveal that the activity of a protein kinase, NIMA, is also required^{3,4}.

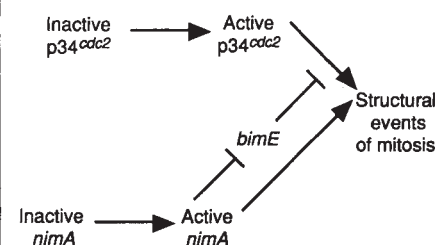
Investigations of the cell cycle in *Aspergillus* were started by Morris, who isolated a large number of temperature-sensitive cell-cycle mutants⁵. These were put into two classes: never-in-mitosis (*nim*) and blocked-in-mitosis (*bim*). Several of the *nim* mutants arrest cells in G2 and are therefore candidates for functions involved in inducing mitosis. Surprisingly, only one of these genes, *nimT*, has turned out to be the homologue of a cell-cycle gene isolated in the budding or fission yeasts (unpublished results quoted in ref. 3); *nimT* is a functional homologue of the fission-yeast *cdc25* and budding-yeast *mih1* genes. The product of these genes is a protein phosphatase that removes phosphate from a tyrosine residue on p34^{cdc2} (refs 6, 7). This phosphate group inhibits protein kinase activity and, at least in

fission yeast, its removal is a rate-limiting step for the entry into mitosis⁸.

Another *nim* gene, *nimA*, is similar in sequence to those encoding protein kinases and is induced as cells enter mitosis. Strong overexpression of *nimA* induces premature entry into mitosis, suggesting that the gene might play an essential part in the induction of mitosis⁹. It has now been shown that the NIMA protein is indeed a protein kinase, and that its kinase activity is regulated during the cell cycle, rising as cells enter mitosis and falling as they leave it⁴. But the most provocative findings arise from comparison of the kinase activities associated with NIMA and p34^{cdc2} in different cell-cycle mutants. Temperature-sensitive *nimA* mutations arrest cells in G2, with little or no NIMA-associated kinase activity, but the level of p34^{cdc2}-associated kinase activity in these cells is as high as it is in mitotic cells. A *nimT* mutation also arrests cells in G2, but has exactly the opposite kinase phenotype (high levels of NIMA-associated kinase activity, but low levels of p34^{cdc2}-associated kinase activity³). So the activation of the two protein kinases proceeds by independent pathways, and both kinases must be activated in order for cells to enter mitosis.

Examination of a third *Aspergillus* mutant, *bimE*, has helped define the role of *nimA* in inducing mitosis; *bimE* is a temperature-sensitive mutant that

enters mitosis from any point in the cell cycle when cells are shifted to the non-permissive temperature¹⁰. The presence of the *nimA* mutation does not block the ability of *bimE* mutants to enter mitosis, from which it would seem that the *bimE* protein acts to prevent entry into mitosis and is normally overcome by the activation of *nimA* (ref. 10). It will be interesting to learn if *bimE* is a substrate for



Interactions between p34^{cdc2}, *nimA* and *bimE* and the induction of mitosis. (—| indicates inhibitory interactions.)

NIMA. Although the mitotic spindle in *bimE* mutants is normal, the spindle in *bimE,nimA* double mutants is abnormal⁴, showing that phosphorylation by NIMA has a structural function in organizing the spindle as well as regulating entry into mitosis. The interactions of p34^{cdc2}, *nimA* and *bimE* in the induction of mitosis are summarized in the figure.

In vitro, p34^{cdc2} can directly catalyse many of the phosphorylations that occur in mitosis (for review see ref. 11). Why then is another kinase required to induce mitosis? One possibility is that *nimA* and *bimE* control the affinity of substrates for active p34^{cdc2} *in vivo*. This possibility is appealing because p34^{cdc2} plays essential roles at two different stages of the cell cycle: Start and mitosis. How does one protein kinase achieve different effects at these two times? One suggestion is that G1 cyclins associate with p34^{cdc2} at Start to direct the kinase to one set of substrates, while the mitotic cyclins that bind p34^{cdc2} at mitosis direct the kinase to another set of substrates.

Two observations run counter to this idea, and beg the question of how cells make Start and mitosis different events: injection of purified complexes of mitotic cyclins and p34^{cdc2} into interphase cells does not induce mitosis¹², and in budding yeast mitotic cyclins can substitute for G1 cyclins in inducing passage through Start (D. Lew and S. I. Reed, personal communication). Perhaps the difference between Start and mitosis reflects the differential modification of the substrates for p34^{cdc2} at different stages in the cell cycle, rather than differences between the targeting capacities of different cyclins. For example, prior phosphorylation by another protein kinase might be required to allow cyclins to bind mitotic substrates and target them for phosphorylation by p34^{cdc2}. In such a

scheme *bimE* would prevent, and *nimA* would stimulate, the phosphorylations required for further phosphorylations by $p34^{cdc2}$. The ability of phosphorylation by one kinase, to make a protein a substrate for a second kinase, has been well documented in other systems¹³.

Do homologues of *nimA* and *bimE* help to regulate mitosis in organisms other than *Aspergillus*? Given the striking functional conservation of other cell-cycle regulators, the answer is likely to be yes. If so, it will reaffirm the notion that studies in many different organisms have helped and will continue to help us identify important regulators of the cell cycle. Finally, the complicated interplay between $p34^{cdc2}$, *nimA* and *bimE* shows that the more we learn about the regulation of the protein kinases that regulate mitosis, the more urgently we need to

understand the details of the interactions between these kinases and their substrates inside the cell. □

Andrew W. Murray is in the Department of Physiology, University of California, San Francisco, California 94143-0444, USA.

1. Nurse, P. *Nature* **344**, 503–508 (1990).
2. Murray, A. W. *Nature* **342**, 14–15 (1989).
3. Osmani, A. H., McGuire, S. L. & Osmani, S. A. *Cell* **67**, 283–291 (1991).
4. Osmani, A. H., O'Donnell, K., Pu, R. T. & Osmani, S. A. *EMBO J.* **10**, 2669–2679 (1991).
5. Morris, N. R. *Genet. Res.* **26**, 237–254 (1976).
6. Strausfeld, U. et al. *Nature* **351**, 242–245 (1991).
7. Kumagai, A. & Dunphy, W. G. *Cell* **64**, 903 (1991).
8. Gould, K. L. & Nurse, P. *Nature* **342**, 39–45 (1989).
9. Osmani, S. A., Pu, R. T. & Morris, N. R. *Cell* **53**, 237–244 (1988).
10. Osmani, S. A., Engle, D. B., Doonan, J. H. & Morris, N. R. *Cell* **52**, 241–251 (1988).
11. Moreno, S. & Nurse, P. *Cell* **61**, 549–551 (1990).
12. Lamb, N. J. C. et al. *Cell* **60**, 151–165 (1990).
13. Fiol, C. J., Wang, A., Roeske, R. W. & Roach, P. J. *J. biol. Chem.* **265**, 6061–6065 (1990).

Inside a black hole

William A. Hiscock

ARE objects falling into a black hole doomed to be crushed to infinite density, or might the interior of a black hole be a 'tunnel' to another universe? Work by Amos Ori¹ shows that the second possibility is not ruled out.

The issue of the final state of gravitational collapse has attracted much attention in theoretical general relativity over the past 30 years. It now seems settled that a sufficiently massive star, at the endpoint of stellar evolution, will undergo gravitational collapse to form a black hole, from which not even light can escape. A major achievement has been the proof of the black-hole uniqueness theorems², which state that any body that collapses to form a black hole will settle down to a final state which may be described by at most three parameters (four if magnetic monopoles exist) — the black hole's mass, its electric charge and its angular momentum. These theorems imply that black holes are in many ways much simpler objects than the stars that collapse to form them, for a complete description of a star would need much more than four numbers.

Unfortunately, the black-hole uniqueness theorems apply only to the portion of the space-time outside the event horizon, the boundary of the black hole. The space-time geometry of a black-hole interior is far less constrained than the final state of the exterior, and most efforts to study the possible state of the interior have centred on the models obtained by simply extending the few unique exterior solutions of the Einstein equations through the event horizon into the interior.

These models reveal several significantly different behaviours. The simplest model, of a 'Schwarzschild' black hole, which possesses mass but no charge or angular momentum, has an all-encompassing gravitational singularity at its core. Any object which falls into such a black hole is inexorably crushed to infinite density at the singularity, where the gravitational tidal forces and deformations become infinite. In contrast, models of black-hole interiors in which the black hole possesses either charge or angular momentum ('Reissner–Nordström' and 'Kerr' black holes, respectively) do not possess such unavoidable singularities. Most matter falling into a black hole described by one of these models will avoid the singularities and emerge from a 'white hole' into another external universe. This would seem to indicate that a black hole can act as a 'tunnel' to other universes. It is perhaps particularly interesting because realistic black holes, formed by the collapse of massive stars, are expected to possess considerable angular momentum (but negligible electric charge).

Beginning in the early 1970s, the accuracy of these simple models came into question. Although the 'tunnel' solutions to the Einstein equations were legitimate, it was unclear whether they were stable. Would a small perturbation to the exact solution, such as starlight falling in from the external Universe, be damped out, or would it grow without bound? A number of mathematical studies (see for example refs 3–5) came to the same conclusion: the 'tunnel' interiors were unstable.

For simplicity, these studies have dealt

with the spherically symmetric Reissner–Nordström black holes, rather than with the less symmetric and more complicated rotating Kerr solutions. The structures of the black-hole interiors are similar, and it is believed that the Reissner–Nordström results apply also in the rotating case. An arbitrarily small external disturbance was found to lead to an exponentially blue-shifted sheet of radiation sweeping across the interior of the black hole. The energy density of the radiation measured by an infalling observer would actually diverge to infinity before the observer could traverse the 'tunnel'. It was widely felt that such behaviour would, in a self-consistent analysis in which the gravitational field of the radiation was taken into account, seal off the 'tunnel', and yield an interior similar to the Schwarzschild model, with an all-encompassing crushing singularity.

More recently, there have been attempts^{6–8} to examine explicitly how the unbounded growth of these disturbances would alter the geometry inside the black hole, in an effort to see if the flux does create an inevitable crushing singularity in the interior. These studies have all made the simplifying assumption that the energy in the disturbance may be approximately treated as a radially ingoing stream of radiation. This allows one explicitly to solve the Einstein equations (the resulting solution is known as Reissner–Nordström–Vaidya space-time) and determine the geometry of the interior. The conclusion is that the stream of infalling radiation does create an all-encompassing singularity, and that the tidal gravitational forces felt by an infalling observer will grow without bound as they approach the singularity.

But this does not quite settle the issue: the singularity in these models is still not as strong as that present in the interior of a Schwarzschild black hole. Although the tidal forces do grow without bound, there also exists a coordinate system in which the curvature of space-time appears to be bounded as the singularity is approached. In the singularity classification scheme of Ellis and Schmidt⁹, this is a "non-scalar curvature singularity" (or, more poetically, as Ellis put it, a "whimper" rather than a "bang"). Such singularities are known, in other contexts, to be generally unstable^{10,11}. The question which then arises is whether additional physical effects, ignored in the simplified models, will strengthen the singularity or remove it.

Poisson and Israel^{7,8} have studied a slightly more complicated and realistic model: they consider the effect of the radially ingoing radiation scattering off the space-time geometry. This will convert some of the ingoing radiation into radially outgoing radiation. They have

not been able to solve the Einstein equations fully because of the added complexity in this case, but they have shown that the general effect of the scattered radiation is to increase greatly the effective mass in the black-hole interior, reducing the effect of any charge or angular momentum and causing the interior to resemble more closely the Schwarzschild interior. They have called this growth of the effective mass, "mass inflation". Mass inflation causes the singularity to become stronger, and the curvature of space-time grows without bound as the singularity is approached. This effect is entirely confined to the interior of the black hole; the mass of the black hole as seen by a distant observer outside the event horizon remains constant and finite.

Enter Ori¹, who has constructed an explicit model which exhibits mass inflation. He was able to solve the Einstein equations by making the simplifying assumption that the scattering of ingoing to outgoing radiation all takes place at one time. In his model, he shows that the space-time curvature does grow without bound as suggested by Poisson and Israel. Similarly, as in the earlier models, the tidal gravitational forces on an infalling observer grow without bound as the singularity is approached. However, Ori also shows that the second integral of the tidal forces, which represents the actual tidal deformation of an infalling body, does not grow without bound. The distortion of a physical object is then finite as it reaches the mass-inflation singularity, and it may be able to travel through the singularity to some further region of space-time.

Ori points out that description of the region beyond the mass-inflation singularity will require an adequate theory of quantum gravity. For the moment, then, we cannot know what might lie beyond. The clear direction for further work is to try to determine if the properties of Ori's exact solution are general, or whether they are perhaps specific to a small set of solutions. □

William A. Hiscock is in the Department of Physics, Montana State University, Bozeman, Montana 59717, USA.

- Ori, A. *Phys. Rev. Lett.* **67**, 789–792 (1991).
- Wald, R. *General Relativity* 322–324 (University of Chicago Press, 1984).
- Simpson, M. & Penrose, R. *Int. J. theor. Phys.* **7**, 183–198 (1973).
- Matzner, R. A., Zamorano, N. & Sandberg, V. D. *Phys. Rev.* **D19**, 2821–2826 (1979).
- Chandrasekhar, S. & Hartle, J. B. *Proc. R. Soc. A* **A384**, 301–315 (1982).
- Hiscock, W. A. *Phys. Lett.* **A63**, 110–112 (1981).
- Poisson, E. & Israel, W. *Phys. Rev. Lett.* **63**, 1663–1666 (1989).
- Poisson, E. & Israel, W. *Phys. Rev.* **D41**, 1796–1809 (1990).
- Ellis, G. F. R. & Schmidt, B. G. *Gen. Relativ. Gravitation* **8**, 915–953 (1977).
- King, A. R. *Phys. Rev.* **D11**, 763–767 (1975).
- Siklos, S. T. C. *Commun. math. Phys.* **58**, 255–272 (1978).

Long-term monsoon regulators

T. Hagelberg and A. C. Mix

As one of the more sensitive components of the weather system, one might expect the Asian monsoon to respond to the ice ages. Yet in a study taking in 350,000 years of climate data, reported on page 720 of this issue¹, Clemens *et al.* find that although the strength of monsoons has varied greatly on long time-scales, changes in ice cover seem to have played little part in this.

The Asian summer monsoon is the product of seasonal heating of the large continental land mass relative to the cooler Indian Ocean. This creates high pressure over the ocean and low pressure

Clemens *et al.* find that the monsoons were particularly strong 9, 50, 123, 215, 261 and 329 thousand years ago (with a resolution of around 5,000 years). But although the data show that the monsoon was relatively weak 18,000 years ago, during the last glacial maximum, a cross-spectral analysis of the variations in each record indicates that, contrary to model predictions³, ice and snow albedo effects are not important in controlling monsoons over long timescales (100,000 years, see their Fig. 5 on page 724).

Why the disagreement with atmospheric general circulation models? Have climate models overplayed the role of ice in controlling monsoon strength? Or do we know too little about the history of ice cover in Tibet to model the albedo effects properly? Other recent studies draw attention to the problems of treating snow and ice feedback in general circulation models. Climate models used to predict future trends in greenhouse warming disagree in the effect of ice⁴: some predict ice feedback to be positive, others indicate it to be negative. With this uncertainty in model feedbacks, the disagreement with geological data is not too surprising.

An alternative addressed by Clemens *et al.* is that low-latitude processes control monsoon intensity. The Indian Ocean data add to the growing consensus that ice-age climates of the tropics are partly decoupled from those of the high latitudes. For example, Atlantic Ocean temperature reconstructions suggest that the equator emphasizes a different (higher dominant frequency) beat than do the poles^{5,6}. And lake-level data link African monsoons to large-scale circulation and heat transport in the Atlantic Ocean⁷.

Clemens *et al.* focus on the role of latent heat in modulating monsoon strength: water vapour carried northwards from the Indian Ocean in the monsoonal winds deposits its latent heat of evaporation into the atmosphere as it condenses and precipitates over land, thus creating a positive feedback⁸. They use the Southern-Hemisphere sea-surface temperature as a proxy for the latent-heat forcing in their observations — increased evaporation (into cold, dry winds) produces stronger cooling of the sea surface. With this measure, they argue that latent-heat forcing is the dominant factor in monsoon strength. Nevertheless, latent-heat extraction is primarily a function of the atmospheric vapour-pressure gradients and wind strength, and although sea-surface temperatures are related, it is difficult to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Monsoon floods, part of a feedback cycle.

ure over land, setting up the characteristic cyclonic summer-monsoon wind patterns. Evaporation from the tropical Indian Ocean adds moisture and latent heat so intensifying the monsoon. Any snow or ice cover over the Asian land mass reflects some of the solar irradiation so reducing the 'sensible' heat source, which (one might expect) would weaken the monsoon. Indeed, snow cover on the Tibetan Plateau is associated with weakened monsoon circulation².

Clemens *et al.* set out to find whether either mechanism operates over the longer timescales represented by the glacial cycles. Their evidence, stretching back 350,000 years, is derived from sediments recovered from two sites, in the Arabian Sea and Indian Ocean. The oxygen isotope (¹⁸O) record of the sediments is their clue to global snow and ice cover and albedo (a proxy for sensible heat); sea-surface temperature (assumed to be a proxy for latent heat) is estimated from the species of foraminifera preserved in the underlying sediments; and monsoon strength is judged from four tracers — the flux of biogenic opal, the abundance of a particular foraminifera that responds to monsoon-induced upwelling in the ocean, the flux of barium and the size of dust grains carried from Somalia and Arabia by monsoonal winds.

discriminate cause and response on long timescales in systems as complicated as this. Coupled models of the atmosphere and ocean, now in their infancy, should be able to explore these relationships⁹.

In their analysis, Clemens *et al.* search in particular for variations at the frequencies of the Milankovich cycles. Changes in the Earth's orbit — in its eccentricity, in the tilt of the Earth's axis and the gradual precession of the perihelion — modify the seasonal and latitudinal distribution of solar radiation, with periods of 100,000, 41,000 and 23,000 years, respectively. The analysis examines correlations and phasing of several monsoon indicators at these periods but it is limited to addressing linear relationships. The appearance of monsoon variations at other frequencies, however, raises the possibility that some monsoon mechanisms may respond in a nonlinear fashion to the external forcing, indicating the operation of feedbacks more complex than previously suspected. Analyses¹⁰ that resolve nonlinear interactions are needed to unravel these mechanisms.

The challenge will then be to compare these more detailed views of the data with model results. Most comparisons so far use general circulation models run to equilibrium. These give great spatial detail for a single moment. In contrast, data from deep-sea cores provide detailed time series of climate history with poor spatial resolution. Systems that involve long time constants, as do ice-sheet feedback and large-scale ocean circulation and chemistry, may violate some of the assumptions of these comparisons. Nonlinear models that simulate palaeoclimatic time series are rare^{11,12}, and although useful they are so highly parameterized that their behaviour can be compared to data only in abstraction. The acquisition, analysis and correlation of geological data, as done by Clemens *et al.*, promise new insights to be built into climate models. □

T. Hagelberg and A. C. Mix are in the College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA.

A birth certificate for CD2

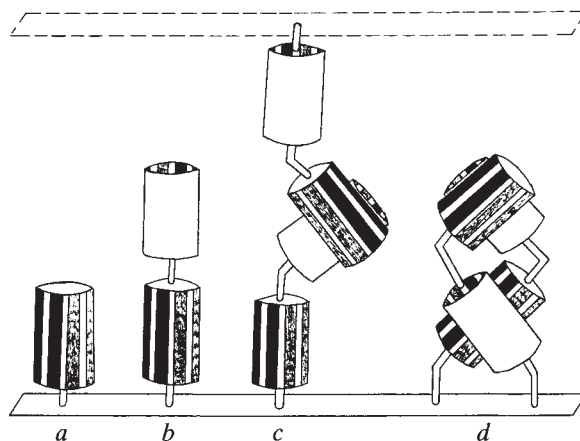
Timothy A. Springer

As the usual tag of 'superfamily' implies, the immunoglobulin-related molecules are a large and extended kindred. They are by far the most populous and functionally diverse family of molecules present on the cell surface. On page 762 of this issue¹, Driscoll and colleagues present a three-dimensional structure for the business end of CD2, an important T-lymphocyte adhesion receptor, and at the same time relieve any doubts about the legitimacy of this molecule as a member of the immunoglobulin (Ig) superfamily.

The CD2 adhesion molecule is one of a number of adhesion molecules on T lymphocytes that participate in interactions with other cells, and hence contribute to antigen-specific responses of both cytotoxic and helper T lymphocytes². It functions by binding to LFA-3, a widely expressed molecule that is also a member of the Ig superfamily. Cell-adhesion and monoclonal antibody blocking experiments, and saturable binding of purified soluble CD2 or LFA-3 to the complementary molecule on the cell surface, have provided convincing evidence that CD2 binds to LFA-3 and measurements of the affinity of this interaction. Water-soluble monomeric molecules give a K_d for binding and a K_i for inhibition of adhesion in the range of 1 μ M (refs 3–6). CD2 contains two Ig-like domains; mutagenesis and proteolysis studies have shown that the N-terminal, Ig-like domain binds to LFA-3 (refs 6,7).

The structure described on page 762 is of this N-terminal, ligand-binding domain in rat CD2. The domain was expressed in *Escherichia coli* and its structure in solution determined by ¹H and ¹⁵N multidimensional heteronuclear NMR spectroscopy. Driscoll *et al.* present high-resolution data on the atoms of the polypeptide chain backbone, and no doubt side-chain data will be forthcoming soon. The polypeptide chain fold is that of an Ig V-region, clearly providing CD2 with an impeccable birth certificate as a true descendant in molecular evolution from the primordial Ig-like domain. Although studies on the binding of CD2

to LFA-3 have to date been conducted on the human and sheep molecules, and rat LFA-3 is yet to be identified, there is every reason to expect that the polypeptide fold will be the same in all species. The polypeptide conformation of CD2 is almost identical to previously solved immunoglobulin V-domains and the CD4 N-terminal V-like domain^{8,9}; the only significant difference is a comparative truncation in β -strands B, D and E, and a compensating lengthening



An evolutionary pathway for dimerization in the immunoglobulin family¹³. a, A primordial, cell-surface Ig domain. The face of the Ig domain with the four-stranded β -sheet is shown with four stripes; the face with the other β -sheet (three-stranded in C-domains or five-stranded in V-domains because of the addition of strands C' and C'') is plain. b, Tandem domain duplication with divergence to give rise to V (upper) and C (lower) domains, as seen in CD2 and LFA-3 (refs 10,15), the first two domains of CD4 (refs 1,8,9), and antibody H and L chains. Rotation by 180° of the V-domain relative to the C-domain is shown as in antibody H and L chains^{13,14}, and in CD4 (refs 8,9). c, Symmetric interaction between the V-domains of two adhesion molecules on opposite cells, as hypothesized for CD2 and LFA-3 and their ancestors. The V-domains are shown interacting exactly as in dimerization of V_L and V_H . d, Dimerization of V_H-C_H1 with V_L-C_L as found in Fab fragments of antibody^{13,14} and hypothesized for the α - and β -subunits of T-cell receptors. The N-terminal (upper) V-domains interact through the five-stranded β -sheets whereas the C-domains interact through the opposite three-stranded β -sheet.

and kinking of the loop between β -strands B and C. Although the cysteines that form the intradomain disulphide loop in most Ig domains are not present in domain 1 of CD2, the homologous residues are the correct distance apart in the structure so that they could form a disulphide bond if they were cysteines.

Prediction of membership of the Ig superfamily is no easy matter because primary sequences have diverged to the extent that a statistically significant relationship often cannot be proven by sequence comparison; hence the importance of three-dimensional structure, which is much more highly conserved. The pre-

1. Clemens, S., Prell, W., Murray, D., Shimmield, G. & Weedon, G. *Nature* **353**, 720–725 (1991).
2. Shukla, J. In *Monsoons* (eds Fein, J. & Stephens, P.) 399–464 (Wiley, New York, 1987).
3. Prell, W. L. & Kutzbach, J. E. *J. geophys. Res.* **92**, 8411–8425 (1987).
4. Cess, R. D. *et al.* *Science* **253**, 888–892 (1991).
5. McIntyre, A. *et al.* *Paleoceanography* **4**, 19–55 (1989).
6. Imbrie, J. *et al.* in *Climate and Geosciences* (eds Berger, A. *et al.*) 121–164 (Kluwer, Amsterdam, 1989).
7. Street-Perrott, F. A. & Perrott, R. A. *Nature* **343**, 607–612 (1990).
8. Hastenrath, S. & Lamb, P. J. *J. phys. Oceanogr.* **10**, 694–708 (1980).
9. Kutzbach, J. & Gallimore, R. G. *J. geophys. Res.* **93**, 803–821 (1988).
10. Hagelberg, T., Pisias, N. & Elgar, S. *Paleoceanography* (in the press).
11. Ghil, M. in *Climate and Geosciences* (eds Berger, A. *et al.*) 211–240 (Kluwer, Amsterdam, 1989).
12. Maasch, K. A. & Saltzman, B. *J. geophys. Res.* **95**, 1955–1963 (1990).

diction that CD2 has an Ig-like structure, based on superfamily sequence patterns, is now clearly vindicated¹⁰; the prediction of secondary structure from primary sequence without taking into account similarities to known structures is fraught with difficulty¹¹, and it is not surprising that this approach failed to yield the correct structure for CD2 (ref. 6).

The new work brings us closer to a molecular definition of the binding site on CD2 for LFA-3. Alignment of the rat and human sequences shows that mutations that disrupt binding to LFA-3 (ref. 7) lie in the region of β -strands C' and F. This leads to the prediction that the face of the β -sheet containing the five β -strands C'', C', C, F and G binds to LFA-3. The structure of CD2 now provides a basis for designing further mutations, particularly in β -strands C'', C and G, that should define the extent of contact between CD2 and LFA-3.

Future structural imaging of CD2 bound to LFA-3, either in a crystal or in solution, would be of great interest. If the general rules for protein-protein association hold¹², the area of contact (calculated on the K_d of 10^{-6} M) will be about 620–680 Å², only slightly lower than the 700–800 Å² involved in the binding of antibodies to protein antigens (K_d 10^{-9} – 10^{-10} M); yet the nature of the contact appears to be quite different.

Whereas in antibodies the contact site between the variable Ig domains and antigen is formed by the three hypervariable loops that connect strands in the β -sheets and are located at one end of the Ig domain, the ligand-binding residues in CD2 can now be inferred to be on the face of a β -sheet. This type of interaction between CD2 and LFA-3 would be more akin to interactions that underlie dimerization of Ig domains in immunoglobulins (see figure). The V-domains of the light and heavy chains of immunoglobulin, for example, interact through the five-stranded β -sheet that forms one face of the domain; known C-domains by contrast turn the other cheek, interacting through the four-stranded β -sheet forming the other face of the domain^{13,14}. Both CD2 and LFA-3 have a single V-like domain N-terminal to a single C-like domain, and it is intriguing that dimerization of V-domains in immunoglobulin uses the same face as in interaction of the V-like domain of CD2 with LFA-3.

The sequence similarity between CD2 and LFA-3 (ref. 15), together with the close linkage of their genes¹⁶, suggests that they arose by tandem gene duplication and this gives rise to a prediction about the symmetry of their interaction. The common ancestor might have participated in like-like (homophilic) interactions with the same molecule on another cell, which evolved to a like-

unlike (heterophilic) interaction between CD2 and LFA-3 after the gene duplication event (c in the figure). The most optimal interaction between identical proteins is symmetric. This symmetry is seen in oligomeric proteins and is preserved within these proteins after gene duplication and divergence of the monomers, as in the pseudo-dyad axes that relate V_H to V_L and C_H to C_L in the antibody Fab fragment (d in the figure). The prediction therefore is that a primordial interaction between the

common ancestors of CD2 and LFA-3 would be preserved in evolution as an interaction between corresponding faces of CD2 and LFA-3 with pseudo twofold rotational symmetry. Whether this prediction is true or false will emerge only when the structure of CD2 bound to LFA-3 is solved. □

Timothy A. Springer is at the Center for Blood Research, Harvard Medical School, 800 Huntington Avenue, Boston, Massachusetts 02115, USA.

1. Driscoll, P. C., Campbell, I. D., Cyster, J. G. & Williams, A. F. *Nature* **353**, 762–765 (1991).
2. Springer, T. A. *Nature* **346**, 425–433 (1990).
3. Selvaraj, P. et al. *Nature* **326**, 400–403 (1987).
4. Dustin, M. L., Olive, D. & Springer, T. A. *J. exp. Med.* **169**, 503–517 (1989).
5. Sayre, P. H., Hussey, R. E., Chang, H. C., Ciardelli, T. L. & Reinherz, E. L. *J. exp. Med.* **169**, 995–1009 (1989).
6. Reeny, M. A., Neidhardt, E. A., Sayre, P. H., Ciardelli, T. L. & Reinherz, E. L. *J. biol. Chem.* **265**, 8542–8549 (1990).
7. Peterson, A. & Seed, B. *Nature* **329**, 842–846 (1987).
8. Wang, J. et al. *Nature* **348**, 411–418 (1990).

9. Ryu, S.-E. et al. *Nature* **348**, 419–425 (1990).
10. Killeen, N., Moessner, R., Arvieux, J., Willis, A. & Williams, A. F. *EMBO J.* **7**, 3087–3091 (1988).
11. Fasman, G. D. *Trends biochem. Sci.* **14**, 295–299 (1989).
12. Chothia, C. & Janin, J. *Nature* **256**, 705–708 (1975).
13. Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M. & Panagiotopoulos, N. *Biochemistry* **14**, 3953–3961 (1975).
14. Amzel, L. M. & Poljak, R. J. *A. Rev. Biochem.* **48**, 961–997 (1979).
15. Seed, B. *Nature* **329**, 840–842 (1987).
16. Sewell, W. A. et al. *Immunogenetics* **28**, 278–282 (1988).

SUN-EARTH CONNECTIONS

The great solar storms of 1989

W. S. Kurth

A SERIES of exceptional solar flares in March 1989, widely reported at the time, had extensive effects on the Earth's ionosphere according to three papers^{1–3} in the *Journal of Geophysical Research*. These effects included the rapid upwards drift of plasma, the strong depletion of ionospheric density at low latitudes (a sort of equatorial ionospheric 'hole') and the creation of wavelike oscillations. These reports are among the first detailing the influence of the most recent maximum in solar activity, the March 1989 events in particular.

The effects of the solar activity on the ionosphere are but an example of the influence of solar activity on the Earth and its environment. A spectacular series of solar X-ray flares was produced by a large, complex sunspot group during the period 6–19 March 1989, causing numerous effects at the Earth which remind us that the Sun's influence on us extends far beyond the radiation that lights our way and warms us. The energetic particles escaping from the active Sun can disrupt communications, upset the routine activities of spacecraft in geostationary orbit, cause crippling power failures, and result in brilliant auroral displays at latitudes much lower than normal⁴.

The March flares produced the "great" magnetic storm of 13–14 March 1989 which, in turn, resulted in many of the manifestations which affect our lives in very direct ways. The exact manner in which solar activity affects the magnetic and plasma environment of the Earth is

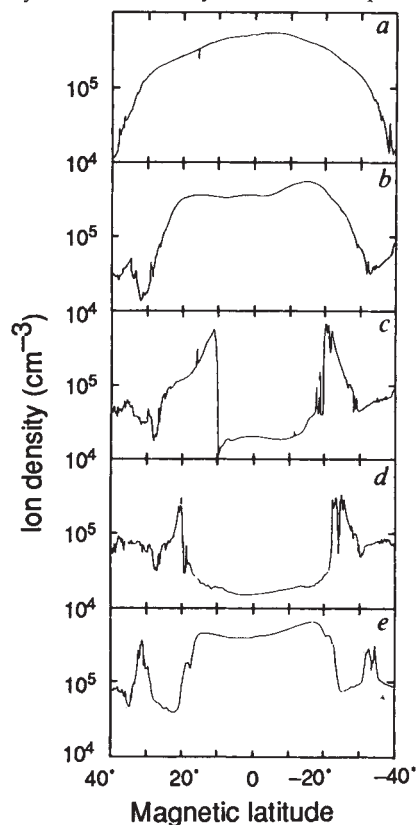
a matter of considerable research. Nevertheless, some general ideas exist for some of the mechanisms. The influx of energetic solar particles results in an increased solar wind pressure at the Earth which, in turn, acts on the magnetosphere. In response to the increased pressure and changes in the electric field due to the solar wind moving past the magnetosphere, the magnetosphere contracts causing such obvious signatures as a decrease in the latitude of the auroral zone from its quiet-time location.

Energetic particles moving Earthward from the geomagnetic tail dump their energy into the atmosphere. When they collide with the upper atmosphere, the atmospheric gases are excited and their subsequent return to a stable state results in the emission of photons which we recognize as the aurora. During active periods, not only are the auroral displays likely to be more intense, but the movement of the auroral oval to lower latitudes means large populations not accustomed to seeing the displays are provided with a rare treat. In the great storm of March 1989, aurorae were observed as far south as Mexico and the Grand Cayman Islands.

The distortion of the magnetic field induces currents in conductors. The power grids at high latitudes are especially subject to these induced currents. The 13 March storm left 6 million people in Quebec without power for up to 9 hours. Power failures occurred at the same time in Sweden.

There are numerous effects observed

on Earth that exhibit an 11-year periodicity, the same as the well-known solar cycle (which is technically a 22-year cycle). These effects include auroral activity, magnetic storms and there are even examples of 11-year cycles in weather patterns⁵. The March 1989 activity marked the beginning of the maximum in solar cycle 22. Extraordinary solar activity was subsequently



Ion density as recorded by the DMSP F9 satellite at 21:20, 23:03, 00:50, 02:32 and 04:12 universal time on the night of 13–14 March 1989. Note the great depletion, or 'hole', in panels *c* and *d*. (From ref. 1.)

observed in September–October 1989⁶ and most recently in March 1991. Gorney⁷ provides a recent review of the effects of solar activity on the Earth, and summarizes several predictions about the current cycle, generally concluding that the maximum will be one of the largest on record. Furthermore, he suggests that periods of high solar activity, like that of March 1989, can be expected for a period of up to 4–5 years.

As discussed by Greenspan *et al.*¹, Batista *et al.*² and Huang and Cheng³, ionospheric disturbances are caused by the heightened activity. The figure, taken from Greenspan *et al.*, shows how ion densities, measured during a series of equatorial passes by a military satellite, decreased dramatically at equatorial latitudes, corresponding to an equatorial 'hole' (panels *c*, *d*). The density variations at midlatitudes also show marked deviations from the normal situation. Greenspan *et al.* suggest that the

penetration of magnetospheric electric fields and the generation of electric fields due to the disturbed motion of circulation patterns in the thermosphere cause large upward drifts of the equatorial plasma. This then causes a poleward drift, which leaves the equatorial ionosphere depleted. As a result of these ionospheric disturbances, the ionosphere's capability to transit or reflect radio waves is disrupted in a number of ways. The resulting communications disruptions affect military and civil operations.

Batista *et al.*² provide ground-based observations from Brazil which support the satellite observations used by Greenspan *et al.* These include measurements of the magnetic-field disturbances, variations in the critical frequency (maximum plasma density) of the ionosphere and the height of the critical frequency, and also the total electron content of the ionosphere. This paper discusses in some detail the role of electric fields in storm-related ionospheric disturbances. It is the penetration of the magnetospheric electric field and the disturbance-dynamo electric fields which are most important in driving the plasma drifts discussed here and with respect to the satellite observations.

Huang and Cheng³ summarize a number of observations made in the Republic of China of the ionospheric disturbances as manifested in magnetic disturbances, total electron content and short-wave propagation effects. Among the most interesting of these are the observations of storm-related enhancements in the nightside total electron content followed by an unusually large decrease. Subsequently, wave-like oscillations were observed simultaneously at three different stations, suggesting "that the whole ionosphere [was] moving up and down during this disturbed period".

Of course, the ionosphere is only one part of the terrestrial system that can respond to solar activity. With such extremes of activity occurring during the current solar cycle, we can look forward to further reports of remarkable reactions in the magnetic field, plasma and neutral gas environment of the Earth. □

W. S. Kurth is in the Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA.

1. Greenspan, M. E., Rasmussen, C. E., Burke, W. J. & Abdu, M. A. *J. geophys. Res.* **96**, 13,931–13,942 (1991).
2. Batista, I. S. *et al. J. geophys. Res.* **96**, 13,943–13,952 (1991).
3. Huang, Y.-N. & Cheng, K. *J. geophys. Res.* **96**, 13,953–13,965 (1991).
4. Allen, J., Sauer, H., Frank, L. & Reiff, P. *Eos* **70**, 1479–1488 (1989).
5. Eddy, J. A. in *The Solar Output and its Variation* (ed. White, O. R.) 51–72 (Colorado Associated University Press, Boulder, 1977).
6. Humble, J. E., Duldig, M. L., Smart, D. F. & Shea, M. A. *Geophys. Res. Lett.* **18**, 737–740 (1991).
7. Gorney, D. J. *Rev. Geophys.* **28**, 315–336 (1990).

Cash on the rail

WITH the collapse of the socialist dream, state industries are everywhere being privatized. Britain may even privatize its railway system. Anyone could then run a train from anywhere to anywhere else whenever they think it might pay, just like a bus or a truck. A separate authority would maintain the track and charge the trains rent. But how could such a chaotic network be signalled and kept safe? Daedalus has an answer.

He points out that two parallel steel rails form a transmission line, an open waveguide along which high-frequency radiation propagates like a sort of guided radar. At any discontinuity, the impedance mismatch reflects a certain proportion of the radiation. By launching a rail-guided signal to front and rear, a train could detect any other train on its track. This cunning rail radar would follow the track round any curve, and could even thread its way through the open point-switches of a complex junction, so as to search only the track that the train will occupy. Trains on other tracks will not interfere.

Radar-guided trains will be able to run safely at unprecedented densities to no central timetable. Each will know the speed and distance of its neighbours at all times. Switches set the wrong way, and obstructions on the line (a detached wagon, say, or a heroine tied across the rails by a moustachioed villain) will also show up clearly. Small track irregularities like level crossings will return weak permanent echoes; these will be mapped throughout the system to serve as navigation markers and 'road signs'.

When a conflict arises — as at a junction, crossover or section of single line — rival trains will negotiate their order of precedence by coded radar pulses. On sound capitalist grounds, Daedalus advocates a direct competitive auction. The switch-point will accept electronic 'bids' from the trains, and will open for the highest bidder. A shrewd train operator would probably give each train a budget (set perhaps by its number of passengers and the fares they have paid) to spend on track rental and setting the points along its route.

Thus the market will be served. Expensive trains full of rich businessmen will bulldoze their way through the system at high speed; cheap outfits carrying pensioners and students will doddle haltingly between them, or make long circuitous detours along empty loops of low-rent track; the track authority will learn how much profit accrues from each specific switch and section of track, so as to extend or abandon it as demand indicates. A fully cost-effective railway will result.

DAVID JONES

Chlorine in the stratosphere

SIR — We have investigated the stable chlorine-isotope composition of several anthropogenic chlorinated organic compounds to test if the compositions of this isotope can be used to quantify the flux of stratospheric inorganic chlorine derived from the decomposition of these compounds. Chlorine has two stable isotopes with an atomic abundance ratio $^{37}\text{Cl}/^{35}\text{Cl}$ of 75.53:24.47 (ref. 1). Chlorine isotope compositions are expressed as $\delta^{37}\text{Cl}$ values in per mil, where

$$\delta^{37}\text{Cl} = \left[\frac{^{37}\text{Cl}/^{35}\text{Cl}_{\text{Sample}}}{^{37}\text{Cl}/^{35}\text{Cl}_{\text{Standard}}} - 1 \right] \times 1,000$$

and $^{37}\text{Cl}/^{35}\text{Cl}_{\text{Standard}}$ is the Cl isotope ratio of standard mean ocean water chloride. The variation in $\delta^{37}\text{Cl}$ values of most naturally occurring Cl compounds is known to be very small^{2,3}, with all measured values between -2 and $+2$ per mil. More important, natural Cl sources available to the atmosphere all have values between -1 and $+1$ per mil. As large kinetic isotope fractionations of chlorine have been observed for processes used to make anthropogenic chlorinated compounds^{4,5}, we expected that each anthropogenic chlorinated compound would have a distinguishable isotope composition.

Chlorine isotope compositions were measured on Freon-12, carbontetrachloride, trichloroethylene, 1,2-dichloroethane, methylchloride and Chlordane (octachloro-4,7-methanotetrahydroindane). Chlorine in these compounds was extracted as LiCl by heating them with lithium metal in a sealed quartz tube. The extracted chlorine was converted into methylchloride and analysed for the isotope composition as described in ref. 6. The standard was prepared by precipitating seawater Cl as AgCl_2 and converting it to methylchloride using the method in ref. 6. Replicate analysis of seawater Cl yielded a value of 0 ± 0.15 (1 σ) per mil for four samples. Several of the methylchloride gas samples analysed for Cl isotope composition were subjected to the whole extraction procedure again to evaluate possible Cl isotope fractionation during the procedure. The original and reprocessed samples yielded identical results.

The analytical results are listed in the table. Methylchloride has a $\delta^{37}\text{Cl}$ value of -6.8 for the vapour and -6.0 for the liquid. All the other compounds analysed have Cl isotope compositions similar to the Cl from naturally occurring Cl-bearing compounds. The depletion of methylchloride in ^{37}Cl clearly indicates preferential incorporation of the lighter chlorine isotope (^{35}Cl) into methylchloride molecules during synthesis under a

large excess of reactant chlorine. On the other hand, given that the kinetic isotope fractionation factor for chlorine (k^{35}/k^{37}) during the synthesis of chlorinated compounds is as large as 1.011 (refs

STABLE CHLORINE ISOTOPE RATIO IN MAN-MADE CHLORINATED COMPOUNDS

Compounds	$\delta^{37}\text{Cl}$ (per mil)
Methylchloride (vapour)	-6.82 ± 0.15 (1)
(liquid)	-6.00 ± 0.14 (1)
1,2-dichloroethane	1.07 ± 0.15 (1)
Carbontetrachloride	0.87 ± 0.15 (1)
Freon-12	1.09 ± 0.15 (2)
Trichloroethylene	2.61 ± 0.15 (2)
Chlordane*	0.19 ± 0.15 (2)

*Octachloro-4,7-methanotetrahydroindane
Numbers in parentheses, number of analyses.

4,5), the results for all other compounds require that chlorine must be quantitatively converted into the products during the synthesis of these compounds by recycling of the unreacted chlorine during the manufacturing process.

These results could allow quantification of the contribution of chlorine from natural methylchloride to the stratospheric inorganic chlorine budget, if natu-

ral methylchloride is also depleted in ^{37}Cl . Methylchloride is the most abundant naturally occurring chlorinated organic compound in the atmosphere; as it has a relatively long residence time there of about 2 years, it is recognized as one of the most important sources of natural chlorine into the stratosphere⁷. Obtaining the chlorine isotope composition of natural methylchloride in the atmosphere may be difficult because its very low volume mixing ratio, as low as 600 parts per 10^{12} (v/v), in the atmosphere and sample size requirements for the analysis do not allow measurements with a gas source mass spectrometer. It should be possible, however, to perform the analyses on a thermal ionization or on an isotope-ratio monitoring⁸ mass spectrometer.

NORIYUKI TANAKA

DANNY M. RYE

Department of Geology & Geophysics,
Yale University, Box 6666,
New Haven, Connecticut 06511, USA

1. Boyd, A. W., Brown, F. & Lounsbury, M. *Can. J. Phys.* **33**, 35 (1955).
2. Hoering, T. C. & Parker, P. L. *Geochim. cosmochim. Acta* **23**, 186–199 (1961).
3. Kaufmann, R. S., Long, A., Bentley, H. & Davis, S. *Nature* **309**, 338–340 (1984).
4. Hill, J. W. & Fry, A. *J. Am. chem. Soc.* **84**, 2763–2769 (1962).
5. Turnquist, C. R., Taylor, J. W., Grimsrud, E. P. & Williams, R. C. *J. Am. chem. Soc.* **95**, 4133–4138 (1973).
6. Taylor, J. W. & Grimsrud, E. P. *Analyt. Chem.* **41**, 805–810 (1969).
7. Cicerone, R. *Rev. geophys. Space Phys.* **19**, 123 (1981).
8. Mathews, D. E. & Hayes, J. M. *Analyt. Chem.* **50**, 1465–1473 (1978).

Bayesian reasoning

SIR — Edwards states¹ that the bayesian claim “that all parametric uncertainty can be measured by probabilities will remain in doubt until compelling arguments are advanced and scientists’ misgivings thus removed”, and opines that “this is probably not possible”, offering references dating back to 1854. Actually, a definitive bayesian argument was published in 1946 by Cox^{2,3}. Apparently the force of this has not been sufficiently appreciated. I shall try to translate the Cox argument into modern idiom, avoiding the technical terms which have failed to convince, and thus try to reduce any misgivings.

Nowadays, I do all my inference with my computer. Its memory has about a billion (10^9) bits, giving 2^{billion} possible patterns. Usually, aided by relevant data, I wish to express my relative preferences for these different patterns.

If I prefer pattern P to pattern Q , and in turn prefer Q to R , then I implicitly prefer P to R . This transitivity (axiom 0) enables me to express my preferences as real numbers $\text{pref}(P)$, $\text{pref}(Q)$,... Though somewhat arbitrary, these are consistently ordered.

The simplest problem has just one bit A , and $2^1 = 2$ patterns, ‘ A ’ and ‘not A ’.

Insofar as I favour ‘ A ’, then I correspondingly disfavour ‘not A ’, and *vice versa*. Hence I have a function f of my numerical preferences such that

$$\text{pref}(\text{not } X) = f(\text{pref}(X)) \text{ (axiom 1)}$$

The next simplest problem has two bits, A and B , and $2^2 = 4$ patterns. To determine my preference for, say, (A, B) with both A and B ON, I simplify. I inspect A first, ignoring B , and assign my $\text{pref}(A)$. If A is OFF, B is irrelevant. If A is ON, then my conditional preference $\text{pref}(B \text{ given } A)$ for B lets me complete my joint preference for (A, B) . Hence I have another function F such that

$$\text{pref}(A, B) = F(\text{pref}(A), \text{pref}(B \text{ given } A)) \text{ (axiom 2)}$$

and this must be the same as $\text{pref}(B, A)$.

The next problem has three bits, say A , B , C . In order to determine my preference for, say, (A, B, C) , I can read the constituent bits in any order, giving six procedures which must all give the same $\text{pref}(A, B, C)$. Cox proved the following. He can always stretch my preferences into revised numerical values for which his functions are $f(x') = 1 - x'$ and $F(x', y') = x'y'$. In other words, he has an absolute scale $\text{prob}(X)$

= stretch(pref(X)) on which his values prob(X) obey prob(A) + prob(not A) = 1, prob(A, B) = prob(A) prob(B given A).

These rules define the only general calculus for patterns containing up to three bits. Larger patterns can be built up by investigating their bits sequentially, during which the rules remain consistent as well as being demanded. Thus a generally applicable calculus can exist, proving existence and uniqueness.

Cox's stretched values are probabilities, and the product rule immediately gives Bayes' theorem, the required tool for inference.

Edwards discards axiom 1, because he has improperly started in an infinite space where he cannot recover a single element by double negation. But in my finite computer, no patterns occupy less than 1 part in 2^{billion} of the space. Negation is the complement operator, which remains reversible for any finite computer: thus $2^{\text{billion}} - (2^{\text{billion}} - 1) = 1$.

Bayesian probability calculus is rock solid and, as Edwards admits, "statistical theory and statistical practice [can thereby] be greatly simplified".

JOHN SKILLING

Dept of Applied Mathematics and
Theoretical Physics,
University of Cambridge,
Cambridge CB3 9EW, UK

1. Edwards, A. W. F. *Nature* **352**, 386–387 (1991).
2. Cox, R. T. *Am. J. Phys.* **14**, 1–13 (1946).
3. Cox, R. T. *The Algebra of Probable Inference* (Johns Hopkins, Baltimore, 1961).

SIR — The example of Fisher's quoted by Edwards¹ is a straw man to discredit the opposing view. Only in this type of argument would a bayesian practitioner confronted with the black mouse give a probability of the mouse being homozygous of one half when s/he knew nothing about any possible relationship between colour and genotype. In practice, a real bayesian practitioner would observe that s/he knew nothing about the genotype and so would assign a uniform prior distribution over the interval 0,1 to the unknown probability. Informally, this means that the bayesian thinks that the probability of the mouse being homozygous lies between 0 and 1, while the knowledgeable scientist believes it is one-half. The seems to be a satisfactory situation, given their respective states of knowledge and belief.

The prior distribution used is subjective, as stated by Howson and Urbach in their Commentary²; however most statistical procedures contain subjective or arbitrary elements (the significance levels for rejecting hypotheses, for example). Here at least the assumptions have to be stated explicitly. In practice, of course, this is not a real difficulty as posterior distributions generally show only weak dependence on the form of

the prior for most reasonable prior distributions when a useful amount of observational or experimental data is available.

It is possible that some non-bayesians (anti-bayesians?) have a conceptual difficulty with the idea of the probability of a probability, which could lead to the above approach being ignored. Fisher's misinterpretation of what to do about ignorance of the values of probabilities in a bayesian argument, is at the root of many arguments against the use of bayesian methods in expert systems, and as we see, this argument is invalid.

RON LARHAM

37 Sancroft Road,
Harrow Weald,
Middlesex HA3 7NU, UK
1. Edwards, A. W. F. *Nature* **352**, 386–387 (1991).
2. Howson, C. & Urbach, P. *Nature* **350**, 371–374 (1991).

Quasar redshifts

SIR — Evidence in support of the cosmological origin of quasar redshifts, which comes mainly from the absorption features in quasar spectra does not seem to have settled the "quasar controversy"¹, as evident from the 'Hypothesis' by Arp *et al.*², who argue that the quasars lie at much smaller distance than their redshift distance, and thus that quasar redshifts are largely intrinsic in origin and not cosmological. However, recent observations on gravitational lensing of quasar images provide clear-cut evidence that quasars lie at their redshift distances.

Einstein's theory of general relativity predicts that a galaxy with a radially symmetric surface density that happens to lie on or very near the line-of-sight to a distant quasar, forms in the sky a ring image³ (Einstein ring) of the quasar with an angular diameter⁴ $\Delta\theta \approx 4\pi(D_{LS}/D_{OS})(v_{\text{cir}}/c)^2$, where D_{OS} and D_{LS} are the angular distances from the observer to the source (quasar) and from the lens (galaxy) to the source, respectively, and v_{cir} is the rotational velocity in the lensing galaxy at the bending point. Such an Einstein ring, MG1654+1306, of a radio lobe of a quasar at redshift $z_s=1.75$ was discovered by Langston *et al.*^{5,6}. It is formed by a bright elliptical galaxy at redshift $z_L=0.254$ and it has an angular diameter of $\Delta\theta=1.97'' \pm 0.04''$. If the quasar and the galaxy lie at their redshift distances then $D_{LS}/D_{OS} \approx 0.73$ and the expected angular diameter of the ring from the estimated⁵ circular velocity in the lensing galaxy ($v_{\text{cir}} \approx 330 \pm 20 \text{ km s}^{-1}$) is $\Delta\theta \approx 2.09'' \pm 0.27''$, in good agreement with the observed value. On the other hand, if the quasar lies in or near the lensing galaxy (that is, if $D_{LS} \leq 50 \text{ kpc}$) then $D_{OS} \approx D_{OL}$, $D_{LS}/D_{OS} \leq 50 \text{ kpc}/D_{OL} \leq 10^{-4}$ and $\Delta\theta$ would be less than 2×10^{-4} arcs, which vastly contradicts the observations.

When the lensing galaxy has an elliptical surface density, the Einstein ring degrades into four images that are located symmetrically along the two principal axes⁷ as observed^{8,9} in the case of Q2237+0305, where the lensing galaxy has a redshift $z_L=0.0394$ and a rotational velocity of $\sim 260 \text{ km s}^{-1}$ (ref. 8), and the quasar images have redshifts $z_s=1.695$. The predicted angular separation between opposite images (roughly the diameter of the ring) is $\Delta\theta \approx 1.85''$, in very good agreement with the ground-based observations^{8,9} ($\Delta\theta=1.75'' \pm 0.10''$), and with recent observations from the Hubble Space Telescope ($\Delta\theta=1.78'' \pm 0.05''$). The system Q2237+0305 was highly advertised as an evidence that a high-redshift quasar lies very near the centre of a low redshift galaxy. However, this would produce an angular separation between opposite images smaller than that observed by at least three orders of magnitude.

One may also compare the observed time delay^{10,11} of 410 ± 10 days between the two images of Q0957+561 at angular distances $|\vec{\theta}_A|=5.24'' \pm 0.5''$ and $|\vec{\theta}_B|=1.00'' \pm 0.5''$ from the centre of the lensing galaxy at redshift $z_L \approx 0.36$ and the predicted time delay¹² $\Delta t_{A,B} \approx 4\pi(1+z_L)(|\vec{\theta}_A| - |\vec{\theta}_B|)(\sigma_l/c)^2(D_{OL}/c)$. A recent measurement of the line-of-sight velocity dispersion in the giant galaxy gave¹³ $\sigma_l = (303 \pm 50) \text{ km s}^{-1}$. Using the best value $H_0 = 67 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for the Hubble parameter, found in a recent review¹⁴, the predicted time delay is 416 ± 130 days, in agreement with the observed time delay. However, if the quasar lies in or near the lens, the deflection angle, and consequently the expected time delay, would be smaller by more than four orders of magnitude than the observed values.

In all other known cases of gravitational lensing of quasar images by galaxies (and clusters of galaxies) the angular separations between the multiple images are of similar magnitude¹⁵. Moreover, the number of lensed quasars that are observed is that expected⁴ if the quasars lie at their redshift distance.

ARNON DAR

Department of Physics,
Israel Institute of Technology,
Technion City, 32000 Haifa, Israel

1. Arp, H. C. *Quasars, Redshifts And Controversies* (Interscience Media, Berkeley 1987).
2. Arp, H. C. *et al. Nature* **346**, 807 (1990).
3. Einstein, A. *Science* **84**, 506 (1936).
4. Turner, E. L., Ostriker J. P. & Gott III, J. R. *Astrophys. J.* **284**, 1 (1984).
5. Langston, G. I. *et al. Astron. J.* **97**, 1283 (1989).
6. Langston, G. I. *et al. Nature* **344**, 43 (1990).
7. Blandford, R. D. *et al. Science* **245**, 824 (1988).
8. Schneider, D. P. *et al. Astron. J.* **95**, 1619 (1988).
9. Yee, H. K. C. *Astron. J.* **95**, 1331 (1988).
10. Vanderriest, C. *et al. Astron. Astrophys.* **215**, 1 (1989).
11. Schild, R. E. *Astrophys. J.* **100**, 1771 (1990).
12. Borgeest, U. *Astrophys. J.* **309**, 467 (1986).
13. Rhee, G. *Nature* (in the press).
14. van den Bergh, S. *Astron. Astrophys. Rev.* **1**, 111 (1990).
15. Blandford, R. D. *Q. J. R. astr. Soc.* **31**, 305 (1990).

Prenatal selection in XO mice

SIR — Ashworth *et al.*¹ attribute to me as a 'personal communication' the statement: "In contrast to humans, mice that are XO are viable with no prenatal lethality." The point I communicated to the authors was that if you compare the preimplantation and postnatal frequency of XOs borne by mice with two X chromosomes, there is no evidence of prenatal selection against XOs. However, there is now good evidence that there is considerable *in utero* selection against XOs borne by mice with a single X chromosome². Because in humans one is dealing with XOs borne by XX mothers, I believe our data provide the relevant comparison.

Why should there be this disparity between the fate of XOs borne by mothers with one as compared to two X chromosomes? I believe the answer lies in the poor quality of the embryos produced by mothers with a single X chromosome, which can be attributed to a deficiency of X-linked products in the developing oocytes³. This poor embryo

quality causes elevated losses during pregnancy, with all genotypes being affected; but I contend that there is preferential loss of the XO conceptuses because they are already developmentally compromised⁴.

PAUL S. BURGOYNE

MRC Mammalian Development Unit,
Wolfson House,
4 Stephenson Way,
London NW1 2HE, UK

1. Ashworth, A. *et al.* *Nature* **351**, 406–408 (1991).
2. Hunt, P. A. *Development* **111**, 1137–1141 (1991).
3. Burgoyne, P. S. & Biggers, J. D. *Dev Biol.* **51**, 109–117 (1976).
4. Burgoyne, P. S. *et al.* *J. Reprod. Fert.* **68**, 387–393 (1983).

RASTAN *ET AL.* REPLY — The point Burgoyne makes was mentioned in our original manuscript, but deleted to comply with the word limit imposed by *Nature*.

SOHAILA RASTAN*

Clinical Research Centre,
Watford Road, Harrow,
Middlesex HA1 3UJ, UK

*Other signatories: Alan Ashworth, Robin Lovell-Badge and Graham Kay.

Mucosal transmission of HIV

SIR — Mucosal transmission of the human immunodeficiency virus (HIV) by the vaginal mucosa in heterosexual, or rectal mucosa in homosexual subjects can be responsible for 70–80 per cent of

cervico-vaginal, oral and rectal epithelia, identifying them by the CD1-specific monoclonal antibody (NA1/34) and their reaction with anti-HLA DR antibodies (see table). We found large numbers of

LANGERHANS CELLS IN EPITHELIAL TISSUES

Tissues	Monoclonal antibodies	
	CD1(NA1/34)	HLA-DR*
Oral epithelium	10/10	10/10
Vaginal epithelium	10/10	10/10
Cervical epithelium		
Ectocervical squamous epithelium	13/13	13/13
Transformation zone	6/7	7/7
Endocervical columnar epithelium	9/13	†
Rectal epithelium	0/14	0/14

Fresh frozen sections of human oral, vaginal, cervical and rectal tissue were stained by monoclonal antibodies using the Avidin–Biotin–peroxidase complex method described in ref. 6. Control sections without the monoclonal antibodies were negative.

* HLA-DR from Becton Dickinson, and from RFDRI, The Royal Free Hospital, London.

† Most endocervical epithelial cells react with anti-HLA DR antibodies, so it is difficult to differentiate those from Langerhans cells.

AIDS. As Barbara Culliton points out¹, heterosexual transmission of AIDS received a great deal of attention at the seventh international conference on AIDS. In particular, William Haseltine's view that AIDS should be recognized as a "lethal venereal disease" needs to be carefully considered. Langerhans cells and the related dendritic cells appear to be the most readily infectable cells by HIV². Indeed, chronic genital or systemic infection with the simian immunodeficiency virus in rhesus monkeys results in significant localization of the virus in Langerhans cells and macrophages of the genital-tract epithelium³.

We investigated Langerhans cells in

Langerhans cells in the vaginal and ectocervical squamous epithelium, and a smaller number in the endocervical columnar epithelium, as well as in the transformation zone of the cervix which could represent the most vulnerable part of the female genital tract in HIV transmission (see table). Oral mucosa contains a comparable number of Langerhans cells to that found in vaginal epithelium but we did not find Langerhans cells in rectal epithelium. Langerhans and dendritic cells may account for HIV infection of the cervico-vaginal epithelium, or after rectal trauma the dendritic cells present in the rectal lymphoid follicles may be infected.

The enhanced risk of HIV transmission via seminal fluid may be accounted for by the high concentration of virus particles (10^8 per ml) in seminal fluid which is much higher than that found in serum⁴. It seems, therefore, that HIV-infected seminal fluid may be transmitted to Langerhans and dendritic cells in the vaginal and oral epithelia. However, there is no convincing evidence that HIV particles in seminal fluid ejaculated during oral sex gain entry through the oral mucosa. Furthermore, there is no real evidence that intimate kissing between seropositive and seronegative subjects may lead to oral transmission of HIV, though this might be accounted for by the low frequency of isolation and low concentration of HIV in saliva.

An alternative mechanism for HIV transmission should be considered and, indeed, antibody mediated enhancement enables HIV to bind CD4-negative cells⁵. We have recently identified Fc receptors for IgG (FcγRIII and FcγRII) in the endocervical and transformation zone epithelia (L. H. *et al.*, in preparation) and in rectal epithelial cells⁶ but not in oral squamous epithelial cells of normal human subjects. We suggest that Fcγ receptors enable HIV-antibody complexes in seminal fluid to bind to cervical or rectal epithelial cells. These cells could then become permissive to HIV infection, or the complex could move to the basal epithelial surface and infect the CD4⁺ T cells, dendritic cells or macrophages in the connective tissue.

It is evident that at least two mechanisms are available for cervico-vaginal and rectal mucosal HIV transmission: direct HIV infection of Langerhans and dendritic cells or the binding of HIV-antibody complexes to Fcγ receptors on endocervical and rectal epithelial cells which may infect these or the adjacent CD4⁺ cells. The two modes of mucosal entry of HIV need to be investigated urgently, as they might be used in cervico-vaginal and rectal routes of immunization against HIV infection.

THOMAS LEHNER

LUMA HUSSAIN

JANE WILSON

MICHAEL CHAPMAN

Division of Immunology, and of
Gynaecology and Obstetrics,
UMDS of Guy's and St Thomas'
Hospitals,
London SE1 9RT, UK

1. Culliton, B. J. *Nature* **352**, 15 (1991).
2. Macatonia, S. E., Lau, R., Patterson, S., Pinching, A. J. & Knight, S. C. *Immunology* **71**, 38–45 (1991).
3. Miller, C. *et al.* 7th Int. AIDS Conf. (Abstr. W. A.1119) **2**, 121 (1991).
4. Borzy, M. S., Connell, R. S. & Kiessling, A. A. *J. AIDS* **1**, 419–424 (1988).
5. Takeda, A., Tuazon, C. U. & Ennis, F. A. *Science* **242**, 580–583 (1988).
6. Hussain, L. A. *et al.* *AIDS* **5**, 1089–1094 (1991).

Protection from X inactivation

SIR — Davies in News and Views¹ draws attention to the growing literature on genes on the X chromosome that are not subject to inactivation, and to the now established finding that such genes are not confined to a single contiguous region extending from the telomere of the short arm. This distribution cannot be accounted for by arrest of a spreading inactivation process. The finding that this class of genes includes some that are also present (although with minor sequence differences) on the Y chromosome suggests a possible solution to the problem of the mechanism of local and distributed protection from inactivation.

A widely accepted explanation² for inactivation of one X chromosome is that inactivation ensures that the dosage of genes is the same in the female (with an XX complement) as in the male (who with an XY complement carries only one copy of X-linked genes). A similar dosage problem can be seen to exist for any Y-linked genes not involved in sex determination. Such genes may be presumed to have an active homologue on the X-chromosome, and to maintain dosage equilibrium they would be expected not to be subject to inactivation in the female.

How could such protection of X-linked Y-homologous genes be achieved? A possible answer is that they might be protected by sequence specific pairing in male meiosis from a process of genomic imprinting. Pairing of X and Y chromosomes in male meiosis is known to extend well beyond the 'pseudoautosomal region' within which there is strict sequence homology and recombination between X and Y (ref. 3). Genes outside this region present only on the X chromosome would be subject to imprinting, whereas those with homologous (although perhaps not identical) sequences on the Y chromosome would be protected.

The hypothesis predicts the following: (1) protection from inactivation will occur when the X chromosome that is preferentially inactivated in the female is the unpaired paternal X. In marsupials and for trophoblast and primitive endoderm (but not for primitive ectoderm and its embryonic derivatives^{4,5}) in the mouse it is the paternal X that is consistently inactivated. However in primitive ectoderm in the mouse, and in man, which X chromosome is inactivated is apparently random. A test of the hypothesis therefore is whether X-Y homologous genes remain active on an inactivated X chromosome that is maternal in origin. (2) Pairing of X-Y homologous genes (for example, ZFY and RPS4Y in Yp11.3 with ZFX in Xp21 (ref. 6) and RPS4X in Xq13 (ref. 7),

respectively) requires an unusual configuration in male meiosis. If such a configuration occurs it might be detected by electron microscopic examination of surface spread synaptonemal complexes.

According to this hypothesis genes on the Y chromosome (other than those that determine sex) are predicted in each case to have a homologue on the X chromosome that is not subject to inactivation. Conversely, genes that are not inactivated on the X chromosome will be expected to have a homologue on the Y. An apparent exception to this rule is A1S9T, a gene that lacks a Y homologue and had been thought⁸ to escape inactivation. It now seems, however, that this gene is susceptible to the normal inactivation process⁹. Therefore, the principle that X-linked genes

that escape inactivation are also represented on the Y chromosome is preserved, and X-Y pairing in male meiosis deserves consideration as a possible mechanism.

TIMOTHY J. CROW

Division of Psychiatry,
Clinical Research Centre,
Northwick Park Hospital,
Harrow, Middlesex HA1 3UJ, UK

1. Davies, K. *Nature* **349**, 15–16 (1991).
2. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* (Springer, Berlin, 1967).
3. Chandley A. C., Goetz, P., Hargreave, T. B., Joseph, A. M. & Speed, R. M. *Cytogenet. Cell Genet.* **38**, 241–247 (1984).
4. Lyon, M. F. & Rastan S. *Differentiation* **26**, 63–67 (1984).
5. Suguwara, O., Takaji, N. & Sasaki, M. *Cytogenet. Cell Genet.* **39**, 210–219 (1985).
6. Leung, W. Y., Lindgren, V., Lau, Y.-F. C. & Yang-Feng, T. *Cytogenet. Cell Genet.* **54**, 151–153 (1990).
7. Fisher, E. M. C. et al. *Cell*, **63**, 1205–1218 (1990).
8. Brown, C. J. & Willard, H. F. *Am J. hum. Genet.* **46**, 273–279 (1990).
9. Zachsenhaus, E. & Sheinin, R. *EMBO J.* **9**, 2923–2929 (1990).

Chitin and nodulation

SIR — In the yeast *Saccharomyces cerevisiae*, the *CSD2* gene is required for chitin synthase III activity and for synthesis of 90% of the cellular chitin¹. A FASTA search of the TRANSGEN protein database using the deduced amino acid sequence of *CSD2* reveals significant similarities to chitin synthases, as expected, to *nodC* proteins from several

(poly-N-acetyl- β -1,4-D-glucosamine), together with the similarity between the *nodC* and *CSD2* gene products, suggest that *nodC* encodes an N-acetylglucosaminyltransferase that synthesizes the oligosaccharide backbone of NodRm-1. The similarity of the *nodC* protein to DG42, a gastrulation-specific protein³, is noteworthy (FASTA

```

      * * * * *
DG42  SVDYVQ-VCD SDTKLDELAT VEMVKVLESN DMYGAVGGDV RILNPYDSF- ----ISFMSS
nodC  SGLDLVL-NVD SDSTIAFDVV SKL-ASKMRD PEVGAVMGQL TASNSGDTW- ----LTKLID
CSD2  FYETVL-MVD ADTKVPFDAL THMVAEMVKD PLIMGLCGET KIANKAQSW- ----VTAIQV
CHS2  LQPTVVTIIVD VGTRLNNTAI YRLWKVFDMD SNVAGAAGQI KTMKGKWLGLK LFNPLVASQN

      * * * * *
DG42  LRYWMAFNVE RACQSYFDCV SCISGPLGMY RNN----- -ILQV FLEAWYRQKF
nodC  MEYWLACNEE RAAQSRFGAV MCCCGPCAMY RRS----- -ALAS LLDQYETQLF
CSD2  FEYIISHHQA KAFESVFGSV TCLPGCFSMY RIKSPKGS DG YWVPVLNPD IVERYSNDVT
CHS2  FEYKISNILD KPLESVFGYI SVLPGLSAY RYRALKNHED GTGPL--RSY FLGETQEGRD

```

```

      * * * * *
DG42  LGTYCT---- LGDDRHLTNR -VLS--MGYR TKYTHKSRAF SETPSLYLRW LNQOTRWTKS
nodC  RGKPSD---- FGEDRHLTI- LMLK--AGFR TEYVPDAIVA TVPDTLKP Y LRQQLRWARS
CSD2  NTLHKNNLLL LGEDRFLSS- LMLKTFPKRK QVFVPAKAC TAPDKFKVL LSQRRRWINS
CHS2  HDVFTAN-MY LAEDRILCWE LVAKRDAKVV LKYPKEATGE TDVPEDVSEF ISQRRRWLNG

```

Alignment of the sequences of the *CHS2*, *CSD2* (GenBank accession number M73697) *nodC* and DG42 proteins. The region displayed exhibits 30% identity between *CHS2* and *CSD2* (11.5 s.d. above random), 32% identity between *CSD2* and *nodC* (6.0 s.d. above random), and 25% identity between *CSD2* and DG42 (8.2 s.d. above random). The program BESTFIT was used. The marks above the sequences indicate the number of identical amino acids: a dot indicates two, an asterisk three, and a bold asterisk four.

species of *Rhizobium* bacteria (initn=280 for *R. meliloti*²), and to the DG42 protein of the amphibian *Xenopus laevis*³ (initn=173). DIAGON comparisons of each of these proteins to the *CSD2* protein identify a common region of ~180 amino acids.

The functions of the *nodC* and DG42 proteins are unknown: *nodC* is required for nodulation in all species of *Rhizobium*⁴ and for synthesis of the extracellular nodulation factor, NodRm-1, a sulphated N-acyl-tri-N-acetyl- β -1,4-D-glucosamine tetrasaccharide⁵. The relationship of this molecule to chitin

initn=342). Could oligosaccharides like NodRm-1 serve as signals during embryogenesis?

CHRISTINE E. BULAWA
WILMA WASCO

Center for Cancer Research,
Massachusetts Institute of Technology,
Cambridge,
Massachusetts 02139, USA.

1. Bulawa, C. E. & Osmond, B. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7424–7428 (1990).
2. Jacobs, T. W., Egelhoff, T. T. & Long, S. R. *J. Bact.* **162**, 469–476 (1985).
3. Rosa, F. et al. *Dev. Biol.* **129**, 114–123 (1988).
4. Long, S. R. & Atkinson, E. M. *Nature* **344**, 712–713 (1990).
5. Lerouge, P. et al. *Nature* **344**, 781–784 (1990).

Delegation of regulation

Margaret Sharp and Simon Shackley

Biotechnology: EEC Policy on the Eve of 1993. *European Study Service, Avenue Paola 43, 1330 Rixensart, Belgium: 1991. Pp. 390. Bfr. 11,800.*

YESTERDAY (23 October) was D-day for European biotechnology — the day when the European Communities' (EC) directives on contained use (EC/90/219) and on the deliberate release into the environment (EC/90/220) of genetically modified organisms first took effect. Their gestation period has been long and, at times, acrimonious; even now much detail still has to be finalized and implementation, the responsibility of member states, will at best be patchy.

Publication of this report on the development of EC policy towards biotechnology is therefore timely. It is a wholly factual account of the emergent regulatory regime, with sections on the three main areas of regulation — contained use, deliberate release and pharmaceuticals. There is also coverage of the main issues in the debate over intellectual property rights, and a detailed summary of the various framework programme initiatives to promote research and training.

An account of this nature dates very quickly. Although issued in 1991, the report is already badly out of date in several important respects and could seriously mislead any manager, administrator or scientist (the audience at whom one assumes it is aimed) relying on it to provide the necessary background information for the implementation of new directives. Indeed, it is difficult to make sense of the current situation, for example the dispute between commission, council and parliament over the regulation of genetically modified pesticides, without knowing something of the fragmentation of responsibilities in Brussels and the shifting sands of the negotiations over regulation as the various directorates general in Brussels battle among themselves.

Omissions

The report is seriously lacking also as a critique of policy. There is no attempt to analyse the rationale of chosen policies, or to cover areas of dispute. As a result, significant issues are ignored — for example, the question of whether regulation should be by product (as is happening in the United States) or by process (as generally happens in Europe). Nor is there discussion of the concept of 'pre-cautionary' regulation developed in Germany, nor of whether regulation should be horizontal (covering all products across sectors) or vertical (with separate

regulations applying, for example, to pharmaceuticals or pesticides). Even more surprising, perhaps, is the absence of any discussion of the infamous 'fourth hurdle', the moves on the part of some directorates general to introduce into the regulatory framework an explicit socio-economic assessment of proven market need.

These criticisms highlight an opportunity missed. What has been achieved after a decade of policy-making? Where does European biotechnology stand in global terms? Where is it heading? And what difference, if any, can community policy make? We will have to look elsewhere for the answers.

Industrialists' answers to these questions are gloomy. They point, rightly, to Europe's lagging competitiveness — poor patenting performance, low public expenditure on research and development, and the flight of capital and trained manpower to the United States. They conclude, predictably but also correctly, that the crucial decisions for the future of biotechnology in Europe lie with private industry, and in particular with the investment decisions of the major companies, which require an environment 'conducive' to such investments.

Three interrelated issues are central to their concerns: public confidence, the regulatory framework and intellectual property rights. Industry recognizes the need for regulation, even in the absence of known hazard, to secure public confidence but, although it prefers the present stance to earlier uncertainties, it tends to take the view that the commission has given way too readily to the environmentalists on regulation and to the traditionalists on intellectual property rights. Industry would like to see Europe adopt the more liberal stance of the United States. Europe's failures in biotechnology, according to industry, are in large part self-inflicted and attributable to these causes.

Obviously, industry is anxious to minimize the impact of regulation (because regulation costs money and effort) and to maximize patent life. The delays and uncertainties created by the lengthy and ill-tempered debates on regulation have undoubtedly deterred investment. But there are times when perhaps industry doth protest too much. The much-quoted example of BASF's decision to locate its biotechnology department in

the United States is a case in point. BASF was already set on a course aimed at internationalizing its research and development locations; given the strength of the US base in the life sciences, the decision to move was logical. Other companies have done exactly the same. It is a moot point in these circumstances whether such overseas investment is 'capital flight' or merely 'globalization'.

One view suggests that it matters little whether a research laboratory is located in Cambridge, Massachusetts or Cambridge, England — or in Munich or Lyon. We all benefit in terms of use and ultimately the profits will be repatriated. Such a view, however, ignores two important factors. First, the need to create and keep high-value-added jobs, and second, the cumulative nature of research and development in these high-technology fields — once the leading edge of research moves elsewhere, it is difficult to regain competence. Europe cannot afford to be cavalier about the loss of research jobs in biotechnology.

Issues

What role, then, for public policy? It is worth mentioning three issues. First, more than other new technologies, biotechnology depends upon science; hence there is a need to foster the research base in the life sciences and industry's links into that base. EC programmes such as BRIDGE can help, but the prime responsibility for supporting the research base still rightly lies with member states. Second, as the technology matures, the crucial decisions are the investment decisions of private industry and the regulatory framework perforce takes centre stage. Although the needs of industry should not always be paramount, the EC are ill-advised to ignore them.

Finally, we all need to acknowledge that biotechnology raises a host of social, political and moral issues which cannot be swept under the carpet — witness the debates that have already taken place over the patenting of life forms and over animal welfare. These issues require constructive and informed debate. The committee being set up to consider the moral and ethical issues associated with the human genome project is an example of good practice. But is it enough? Should not the EC consider strengthening and extending the European parliament's small Office of Technology Assessment (OTA) to give the EC an institution comparable to the OTA in the United States? □

Margaret Sharp and Simon Shackley are at The Science Policy Research Unit, University of Sussex, Falmer, Brighton BN1 9RF, UK.

A shot in the dark

Fekri A. Hassan

Centuries of Darkness. By Peter James. Jonathan Cape: 1991. Pp. 434. £18.

TIME has puzzled and intrigued thinkers for countless generations. For more than a century, historians and archaeologists have laboured to establish a credible chronology. Their passionately held notions of historical developments have been challenged by new discoveries and by contrary notions from colleagues working in other regions or motivated by different goals. The result has been a procession of new visions and revisions. In *Centuries of Darkness*, Peter James, in collaboration with four contributors (I. J. Thorpe, Nikos Kokkinos, Robert Morkot and John Frankish), presents a challenge to what has emerged in recent years as a synthesis of conventional views on world chronology from 1200 to 800 BC, the synthesis being the best of all possible worlds of controversial and contradictory data. The book may be disquieting for those who cherish the comfort of familiar schemes or find in conventional chronology a justification for their favourite or inherited notions of history. But to suggest, as James and his collaborators do, that an older generation of historians have closed minds is hardly a token of the veracity of this newer generation's claims. Their sensational tone and the ardency of their advocacy besmirches their scholarly

efforts and poisons the air of reasoned scholarly exchanges. This is regrettable because James and his collaborators are professional archaeologists who raise many points worthy of critical examination.

The book purports to examine the chronology of numerous Mediterranean cultures (Italy, Greece, Cyprus, Crete, Palestine, Egypt and Nubia, Mesopotamia (now eastern Syria and Iraq) and Iran) to show that scholars have exaggerated the duration of a period referred to as the Dark Age in Greece. This period witnessed the Trojan wars, the expansion of Arameans, the age of the Ramessides, the transition to the Iron Age, the maritime commercial ventures of the Phoenicians and the spread of the alphabet. James and his collaborators work their way from one culture to another with a sledge-hammer, attacking any chronological scheme that does not fit their grand thesis. Amidst the rubble, they reach Egypt, which they regard as the benchmark for other chronological speculations.

We are thus bound to pay special attention to their proposal of a new Egyptian chronology. Doubts are indeed warranted in many cases. Nevertheless, they have not succeeded in hewing a foundation for an alternative firm chronology. In addition, their presentation of the calibrated radiocarbon dating for Egyptian eighteenth and nineteenth dynasties, curiously relegated to a footnote, does not instill confidence. Contrary to their claims, the calibrated radiocarbon dates from the Ramesside period (including an estimate of $1,230 \pm 96$ calendar years for Ramesses II) are

highly congruent with the conventional chronology for that time interval. Moreover, these dates are bolstered by the best radiocarbon age estimate in the Egyptian sequence — a series of highly consistent age measurements from Tell el-Amarna ($1,333 \pm 50$ calendar years BC). This firm date provides a basis for accepting a conventional chronology for the Hittites and the Assyrians. It lends credence to the conventional correlation of Assurballit, king of Assyria, mentioned in the Amarna letters, with Assurballit I, contrary to the claims set out in the book. Also, if, as widely accepted, the interval between the Late Bronze and Early Iron Age in the Mediterranean is contemporaneous with the late nineteenth and early twentieth dynasties of Egypt, the former would date to about 1200 BC as in the conventional chronology.

James and his colleagues are to be congratulated for the breadth of their treatment, but the reader should not expect new chronological findings or a methodological breakthrough. There is also no critical evaluation of traditional methods of dating using pottery style. Unfortunately, we still face the same chronological quagmire. The vagaries of the archaeological record and the subversion of chronological indicators by cultural processes (for example, the 'Doppler' effect in ceramics), and occasionally archaeologists themselves, have undermined the most ingenious chronological schemes. In addition, precise dates provided by classical writers appear deceptively accurate. Consequently, chronological estimates are inherently highly probabilistic and hence

Revealing the mystery of the Aegean



THE volcanic Santorini islands have fascinated scholars from the great Greek historian Herodotus to Duke Jacomo I Crispo of Naxos, the first to take a scientific interest in the islands. Study of the area has deepened over the past few decades thanks largely to three international congresses held to study the islands. The collected papers of the second congress, from which this fresco detail is taken, were published in 1978, and now three volumes of

the third conference proceedings, *Thera and the Aegean World*, have recently been published by the Thera Foundation. Examining everything from Thera art, technology and volcanology to the 'Theran event and its global impact', these latest 'instalments' offer a wider analysis of this remarkable region, equated by some with the lost city of Atlantis. Further information available from the Thera Foundation, 105-9 Bishopsgate, London EC2M 3UQ, UK.

amenable to preferential interpretations. So it is not surprising that the chronological data for Mesopotamia and Egypt were used by J. Mellaart in 1979 in *Antiquity* to support a high chronology.

Luckily, advanced chronometric methods provide an alternative framework for judging historical data. The current possibility of obtaining precise and accurate radiocarbon dating, coupled with the potential for dating minute samples, promises to provide the type of radiocarbon chronometry that can be matched with the wiggles in the calibration curve as proposed by Colin Renfrew in the foreword to the book. The chronometry of Egypt must be given a high priority as the centrepiece in the puzzle. It is essential to obtain further reliable determinations for the eighteenth to twenty-sixth dynasties, as well as for the twelfth dynasty.

In the past, a lack of appreciation of the importance of radiocarbon measurements, methodological problems and the indiscriminate use of radiocarbon age estimates have confounded rather than clarified chronological problems. Although several methodological problems have been overcome during the past decade, many archaeologists are still sceptical of the usefulness of radiocarbon dating, a scepticism that hinders developments and perpetuates the current chronological impasse.

Just who will read the book? Its popular cast may attract the public, but its technical details will probably reduce its general appeal. If it were written for the specialist, then a scholarly monograph would have sufficed. The book really belongs to a genre of discourse that appeals to professionals who are not actively involved in chronological studies. Given the nature of existing historical data and the prevailing modes of prejudicial reasoning, nonexperts must suspend judgement. They should also be wary of the hyperbolic claims on the dust jacket. I find the book disquieting not because of the doubts it casts on shaky chronologies, but because it confuses scholarship with the excessive zeal and self-righteousness of revolutionary advocacy.

Centuries of Darkness is a worthwhile contribution because it highlights the sticky chronological problems that must be resolved by future excavations and determinations of chronometric age. If it succeeds in stimulating a systematic and vigorous investigation of existing chronologies with a long-due consideration for the pitfalls of historical and stylistic crossdating, it may well prove to be more than a flash in the pan. □

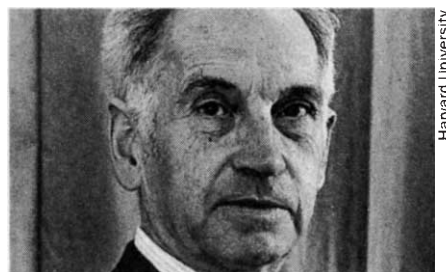
Fekri A. Hassan is in the Department of Anthropology, Washington State University, Pullman, Washington 99164-4910, USA.

Grand masters, great debates

Peter J. Bowler

One Long Argument: Charles Darwin and the Genesis of Evolutionary Thought. By Ernst Mayr. *Harvard University Press: 1991. Pp. 187. \$19.95. (To be published in Europe by Viking on 30 January 1992. £17.99).*

SUDDENLY there is a flood of books on Charles Darwin, after years in which the 'Darwin industry' has confined itself to the production of technical articles. Although there is no consensus yet over



Ernst Mayr — 'concise and elegant'.

a new interpretation of Darwin, many academics now feel confident enough to risk book-length surveys of his life and influence. Ernst Mayr is one of these. As a biologist interested in the history of evolution theory, he now offers us not a biography but his "mature reflections on Darwin's thought" in the form of a history of darwinism. Those familiar with his monumental *Growth of Biological Thought* (Harvard University Press, 1982) will find additional material on Darwin and darwinism here, culled from Mayr's own articles and supplemented by several specially written chapters.

Much of the book is concerned with the development, content and reception of Darwin's theory — or theories, because Darwin described the *Origin of Species* as "one long argument" for a group of linked but partially independent propositions. As if to reinforce this last point, the main chapter on the development of natural selection comes after the discussion of the *Origin's* reception. In his desire to present Darwin as the founder of the most important trend in evolutionary thought, Mayr offers us a very 'modern' Darwin. He concedes some backsliding in Darwin's views on the nature and origin of species, but emphasizes the generally radical character of Darwin's thinking when measured by the standards of his time.

There is, in fact, a good deal of disagreement about Darwin's opinions on religion and the progressive character of evolution. Mayr sides with those who see Darwin as an early agnostic who

kept on using teleological language to satisfy public opinion as well as to appease his wife. Mayr also insists that Darwin's view of evolution was not an expression of the Victorian philosophy of 'progress through struggle' expounded by Herbert Spencer; indeed it threatened the very foundations of this new kind of teleology. Here Mayr runs counter to the many historians who now accept that Darwin did not escape the progressionism of his time. In their efforts to resist the 'modernized' image of Darwin offered by Mayr and other biologists, some revisionists may now be pushing too far in the opposite direction; nonetheless, Mayr's interpretation is controversial. In *The Meaning of Evolution* (forthcoming from the University of Chicago Press), Robert Richard offers a rival image of Darwin that places him firmly in the progressionist camp.

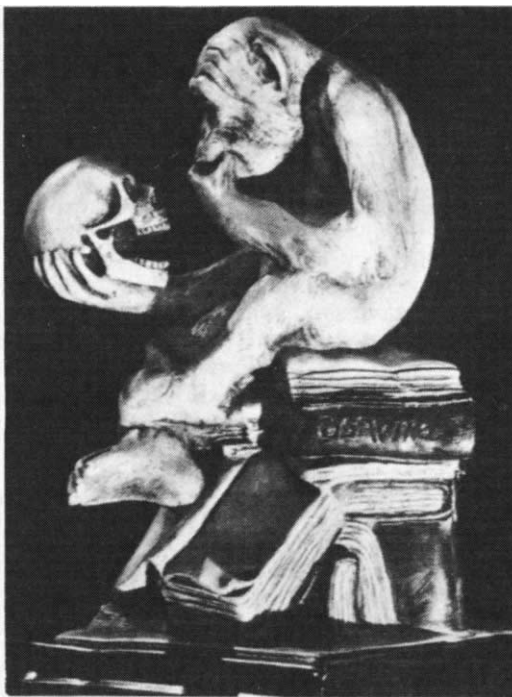
In analysing Darwin's path to natural selection, Mayr takes another controversial line by emphasizing the role of animal breeders in preparing Darwin to interpret Thomas Malthus's famous *Essay on the Principle of Population* (1798) in terms of individual struggle. (Malthus's principle that excessive fertility in nature produces competition for survival gave Darwin and Wallace the foundation they needed for the theory of natural selection.) Many now accept that Darwin's analogy between artificial and natural selection was a product of hindsight; much of the breeder's literature was not, in any case, concerned with selection. But Mayr's real purpose is to use history to display the inner logic of Darwin's arguments, and he concedes that this approach favours an interpretation that may not coincide with the analysis of Darwin's research notebooks.

Mayr goes on to emphasize the radical character of Darwin's thinking by noting how few of his contemporaries were able to follow him towards a completely selectionist view of evolution. He suggests that the real purpose of the *Origin of Species* was to make the case for the fact of evolution and the theory of common descent. Natural selection, the main mechanism of evolution that Darwin proposed, was marginal to the campaign because Darwin knew that it would be controversial — indeed, some of his chief supporters, T. H. Huxley included, did not accept it. Mayr is again in disagreement with other historians by seeing the early form of darwinism as unified around a commitment to natural evolution and common descent (although he does acknowledge the influence of a few borderline figures such as Asa Gray who continued to hope for divine control of evolution). The eventual emergence of natural selection as the dominant theory of the evolutionary mechanism in the 1930s

and 1940s is now described as a 'second darwinian revolution'.

Mayr's emphasis on the nondarwinian character of most late-nineteenth-century evolutionism is important, although we can perhaps forgive him for spending more time on those who did accept natural selection (for example, August Weismann) than on those who did not. He is surely correct in his claim that much of the enthusiasm for nondarwinian mechanisms such as lamarckism and saltationism was generated by a desire to retain as much as possible of the old philosophies of teleology and essentialism.

Mayr's claim that Darwin was successful in promoting the theory of common descent is not as clear-cut as he would have us believe. The early darwinians accepted that 'related' species possess similar characteristics because they are descended from a common ancestor. Most lamarckians and supporters of orthogenesis (directed evolution) challenged this interpretation of 'relationship'. They invoked both adaptive convergence and nonadaptive trends to argue that similar characters can be developed indepen-



Thinking it through.

dently in widely separated lines of descent. Although proponents of this idea did not deny that the lines linked up if they are traced far enough back into the

past, their approach undermined the value of the theory of common descent for tracing phylogenies. This view of evolution was also part of a campaign to retain more traditional values, as can be seen in the efforts made by many early palaeoanthropologists to distance the human race from the apes. By concentrating on ideas about the evolutionary mechanism, Mayr has missed an important point that can be revealed only by looking at how the idea of evolution was actually used to reconstruct the history of life on Earth.

Mayr ends by outlining his views on the most recent developments in evolutionary theory. He is convinced that important steps have yet to be taken, but equally convinced that these steps will take off from the foundation already established. Mayr can hardly be counted as an impartial judge of darwinism's role in the development of evolutionism, but it is valuable to have his "mature reflections" expressed so concisely and elegantly. □

Peter J. Bowler is in the Department of Social Anthropology, The Queen's University of Belfast, Belfast BT7 1NN, UK.

Meddling with intent

John Durant

The Art of Scientific Rhetoric. Edited by Marcello Pera and William R. Shea. *Science History Publications/USA, Box 493, Canton, Massachusetts 02021. 1991. Pp. 212. \$39.95.*

It has recently become fashionable to subject scientific texts to rhetorical analysis. To many scientists, this seems subversive. As Gerald Holton points out in his contribution to this volume of essays, practitioners of modern science claim to use an 'objective' method of demonstrating how nature works; and they oppose in principle all modes of persuasion that rely on subjective factors such as the personal characteristics of individual scientists. Thus, in a draft preamble to the original statutes of the Royal Society of London, Robert Hooke asserted that scientists did not intend to 'meddle' with 'rhetoric'.

What, then, is the point of talking about scientific rhetoric? For some analysts, the point is precisely to prick the bubble of scientific objectivity. For them, the demonstration that science is rhetorical is taken to show that science is not a strictly rational enterprise, or even that science provides no genuinely objective knowledge of the world at all. On

this view, scientific truth is exposed as merely the sum of all those things that scientists have persuaded themselves to believe.

This is obviously silly. If rhetoric and logic are set up as rival explanations for the vast corpus of scientific knowledge, then logic wins easily. Scientific knowledge may be partial, and it may be provisional; but it certainly is not an arbitrary rag-bag of all the propositions that a professional community of scientific rhetoricians has taken to its bosom.

Instead of opposing rhetoric with logic, most of the essays in this volume take the far more sensible view that the two play complementary roles in science. In the words of the editors, "The central question . . . is not whether science gives us genuine knowledge but how it justifies the knowledge that it acquires, transforms and disseminates".

There are two related features of science that make rhetoric an essential part of this process of justification. First, science is not produced solely in a two-way encounter between nature and the individual scientist; rather, it is produced in a three-way encounter between nature, the individual scientist and the community of research colleagues. Scientists are obliged to convince their professional colleagues of the validity of their findings; this obliges them to provide written accounts of what they do.

Second, there is no completely rigorous system of scientific thought. In other words, there is no formal algorithm

whereby evidence may be translated into discoveries. In Karl Popper's terms, science is conjectural; and this means that accounts of scientific research contain more than syllogisms. Typically, what they contain are arguments in favour of particular interpretations of the available evidence.

The twin facts that science involves obtaining the assent of others and that it is conjectural make persuasion inescapable. As Richard Westfall observes, there is a world of difference between Galileo's scornful invective against the scholastics and Newton's icy disdain for both his forebears and his critics; but in their different ways, each man defined, addressed and won over a particular audience to their point of view. Each man was skilled in both logic and rhetoric. The essays in this volume provide a useful guide to the rhetorical devices that scientists have found useful in their work.

Sir Peter Medawar once famously observed that we should not be deceived by the strict conventions of the scientific paper. Beneath the calm surface, there are deep currents of rhetorical intent. This fact should not be taken as evidence that science is imperfect, or that its claims to objectivity are bogus. Rather, it should be taken as an important aspect of the enterprise by which we gain a better understanding of the world. □

John Durant is at the Science Museum Library, Exhibition Road, London SW7 5NH, UK.

A structural taxonomy of DNA-binding domains

Stephen C. Harrison

The structures of several classes of DNA-binding domains reveal a variety of designs for recognizing a specific site on DNA.

MANY DNA-binding proteins recognize specific sites through small, discrete domains. In some cases these domains can be interchanged between proteins, showing that they are independently folded units. Several groups have been identified on the basis of primary sequence relationships, and three-dimensional structures have been determined for members of many of these groups. Here I present a brief illustrated taxonomy of those DNA-binding domains for which structural data are available (see Table 1), and outline how they bind to specific sites. I use the word 'domain' to refer to a protein fragment that can fold correctly without the rest of the polypeptide chain and the word 'motif' to refer to a conserved substructure that does not fold independently.

The domains described here reveal a variety of different designs. They all fold in such a way that they present a protruding surface, a flexibly extended structure, or both, in order to contact DNA base pairs. In most cases, these contacts occur in the major groove. They include hydrogen bonds, both direct and water-mediated, and non-polar van der Waals interactions. An α helix is a frequently occurring but not universal recognition element. The structures show a variety of ways for inserting an α helix in the major groove of DNA. Interactions with atoms of the sugar-phosphate backbone—especially hydrogen bonds to non-esterified phosphate oxygens—are critical for positioning the domains so that their protruding structures align correctly with the base pairs they contact. They are also important for fixing the DNA in a particular conformation. Thus, sugar-phosphate interactions are part of specific recognition, which is a property of the entire protein/DNA interface.

The helix–turn–helix

Members of the most thoroughly studied group of DNA-binding proteins contain domains with a helix–turn–helix (HTH) motif^{1,2}. Most of the prokaryotic transcriptional regulatory proteins fall into this group, as do the eukaryotic homeodomains. An important difference between prokaryotic HTH proteins and homeodomains is that the latter are able to interact with their sites as monomers, whereas most of the former bind as dimers. The HTH is generally defined as a 20-residue segment, with two α helices that cross at an angle of about 120°. There may be additional helical residues, at the N terminus of the first helix or the C terminus of the second, which extend beyond the central 20-residue 'core.' The HTH motif is found embedded in domains of remarkably varied structure². Some are illustrated in Fig. 1a–d (refs 3–8). In all cases, one or more additional elements of secondary structure are needed to enclose the hydrophobic core of a complete domain. In two different all-helix bundles, *trp* repressor⁹ and the inversion stimulation factor Fis (ref. 10), helices from the two subunits of a dimer interdigitate, and the isolated subunits may not fold very stably by themselves.

The second helix in the HTH has been called a 'recognition helix,' because it lies in the major groove when HTH-containing domains bind DNA and because experiments that involve changing residues on its outer surface suggest that it is critical for specificity¹¹. The presentation of this helix to the edges of base pairs exposed in the groove depends on the structure of

TABLE 1 DNA-binding domains for which there are published structural data

Protein	Structural data*	Ref.
Group I (HTH)		
Eukaryotic regulatory proteins		
Homeodomains		
<i>Antp</i>	NMR (†)	3, 12
<i>engrailed</i>	X-ray†	4
Prokaryotic regulatory proteins		
<i>lac</i> repressor	NMR (†)	5
λ repressor	X-ray†	6, 71
434 repressor	X-ray†	7, 72
434 Cro	X-ray†	73, 74
λ Cro	X-ray (†)	75, 76
catabolite activator protein (CAP)	X-ray†	8, 77
<i>trp</i> repressor	X-ray†	9, 78
FIS (inversion stimulation factor)	X-ray	10
Group II (Zn-binding)		
Class 1 (Zn fingers)		
Xfin (finger 31)	NMR	22
ADR1 (finger ADR1b)	NMR	23
Zif 268 (fingers 1–3)	X-ray†	24
Class 2 (LZnH)		
glucocorticoid receptor	NMR; X-ray†	26, 28
oestrogen receptor	NMR	27
Class 3		
GAL4	(NMR)	19
Group III (bCC or bZip)		
GCN4	(NMR) (coiled-coil only)	43
Group IV (β ribbon)		
Class 1		
Met J	X-ray†	58, 59
Arc	NMR	60
Class 2		
HU	X-ray	62, 63

* Coordinates from X-ray crystallography or NMR. The symbol † indicates that the structure of a DNA complex has been published. Parentheses indicate results of lower resolution, as in the NMR structures of *lac* and *Antp* DNA complexes and in the crystal structure of the λ Cro complex, or partial results, as in the NMR studies of GAL4.

the rest of the domain and on the way it contacts DNA backbone. In some cases, such as the N-terminal arms of homeodomains and of lambda repressor, additional elements make direct base-pair contacts. The conformation of the DNA site that is established when the protein binds can also influence the geometry of the major groove itself and hence affect the HTH/DNA interaction².

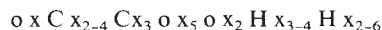
DNA complexes of 434 repressor⁷² and of the *engrailed* homeodomain⁴ are compared in Fig. 2a, b. The 434 repressor binding domain is anchored to segments of DNA backbone on both sides of the major groove, largely through hydrogen bonds to phosphates. These 'positioning contacts' determine the alignment of residues in the recognition helix with particular base

pairs in the operator¹³. There is no simple code that relates operator sequence to a three-dimensional pattern of amino-acid side chains, because in general several side chains may interact with a given base pair, and more than one base pair can be in contact with a given side chain. If two proteins have very similar positioning contacts, however, then appropriately aligned base pairs interact with corresponding residues of the recognition helix, and definite regularities can be seen. For example, there is a 'recognition pattern' common to 434 and lambda repressors, involving glutamines at the N-termini of both HTH helices and an adenine near the outside of the operators¹³. The positioning of the homeodomain HTH on DNA, as seen in two crystal structures (ref. 4; C. Wolberger and C. O. Pabo, personal communication) and by NMR in solution¹², is quite different from any found in complexes of prokaryotic HTH proteins. The entire motif is shifted and reoriented with respect to the DNA. There are key positioning contacts to phosphates on the DNA backbone segments that flank the place where helix 3 lies in the major groove. The N-terminal arm extends into the minor groove to make important specific interactions. Other proteins seem to enhance specificity by helping to position, or reposition, the homeodomain itself. An example is the cooperative binding of yeast $\alpha 2$ and MCM1 (ref. 14). Therefore the diagram in Fig. 2a may not represent a universal homeodomain interaction, or even the only possible interaction for *engrailed* itself. The essential similarity of the *engrailed* and yeast $\alpha 2$ co-crystal structures (ref. 4; C. Wolberger and C. O. Pabo, personal communication) and the *Antp* complex worked out by NMR^{3,12} does, however, suggest that there is a common binding mode in the absence of other factors. Experiments that interchange appropriate amino-acid residues between *bicoid*- and *Antp*-like proteins indicate that the ninth recognition-helix residue has a role in specifying the base 'X' of a TAATX binding-site¹⁵. The overall orientation of the homeodomain shown in Fig. 2a is consistent with this result.

Zn-binding domains

The second group of DNA-binding domains contain Zn as a structural element. There are three classes so far recognized, each with a different structure. The first is the original 'Zn-finger,' an approximately 30-residue module with one Zn ion liganded by two cysteines and two histidines^{16,17}. The second is an approximately 70-residue domain found in the receptors for the steroids and related hormone-like molecules, with two Zn ions, each liganded by four cysteines¹⁸. The third is found in a set of yeast activators, including GAL4, with two closely spaced Zn ions sharing six cysteines¹⁹. I will refer to these as Zn domains of classes 1, 2 and 3, respectively.

The sequences of over 200 class-1 Zn domains (Zn-fingers) conform to the following consensus:



where o is a hydrophobic amino acid and x is variable in character; the invariant Cys and His residues can be taken as a definition of this structure²⁰. Most proteins containing these modules have three or more fingers in direct succession. The folded structure (Fig. 1e) is a compact unit with a 12-residue helix packed against a β hairpin²¹⁻²⁴. The two cysteines flank the turn in the hairpin, and the two Zn-liganding histidines are on the inward-facing side of the helix. The direct concatenation of several fingers in class-1 Zn-binding proteins suggests that they form a repeating structure when bound to DNA²¹. The structure of a three-finger segment from Zif 268 bound to cognate DNA shows just such a repeat²⁴. The fingers bind in essentially identical manner, at intervals of three base pairs (bp). They lie in the major groove, with the N terminus of the helix closest to the edges of bases (Fig. 2c). The ends of the polypeptide chain, as they emerge from each Zn-coordinating module, project in such a way that the protein can wrap around the DNA. Each finger contacts three base pairs, and there are no gaps between

the 3-bp sites seen by adjacent fingers in the protein. Zif 268 is very similar to Krox-20 and Sp1, two proteins that recognize related consensus elements. Interchanging fingers between Krox-20 and Sp1 creates interchanged specificities consistent with the 3-bp repeat seen in the Zif 268 crystal structure²⁵. Most of the interactions in the Zif 268-DNA complex are with one DNA strand. The amino-to-carboxy-terminal order of the fingers corresponds to the 3' to 5' direction of this strand.

Some of the detailed interactions seen in the Zif 268-DNA complex may reflect common features of Zn-finger binding. There are contacts between arginine and serine side chains and DNA phosphates, and there is a conserved hydrogen bond from the first Zn-coordinating histidine in each finger to a DNA phosphate. The interaction made by the histidine is noteworthy, because this residue is an invariant Zn ligand and hence an invariant structural element. Its participation in a key positioning contact suggests that most class-2 Zn domains may bind to DNA in the same way. Base-pair contacts are made by the residue immediately preceding the α helix and by residues 2, 3 and 6 in the helix itself. An interesting recognition pattern involves the residue N-proximal to the helix, which is Arg in all three fingers, and residue 2 of the helix, which is Asp. A hydrogen bond/salt bridge between these side chains holds the arginine in a conformation appropriate for donating hydrogen bonds to N7 and O6 of a guanine.

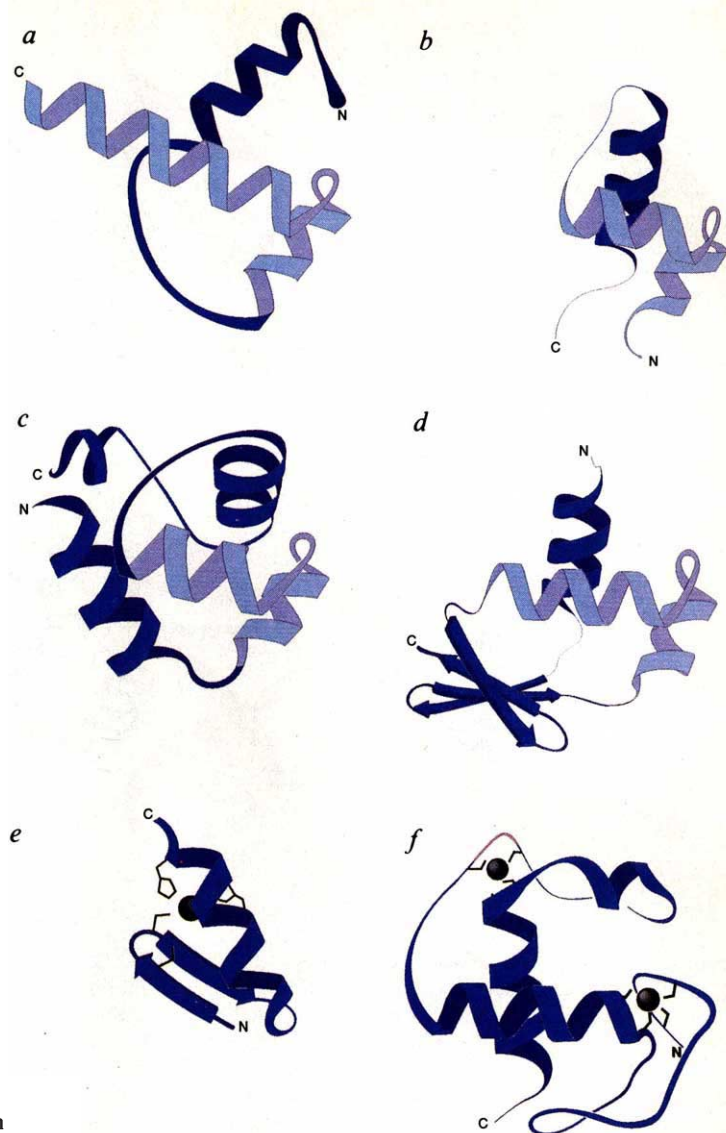
The structures of two class-2 Zn domains have been determined by NMR (Fig. 1f)^{26,27}. They contain two loop-helix elements, each with a Zn ion liganded by two cysteines at the beginning of the loop and two others at the N-terminus of the α helix. The two helices in the domain are packed against each other at nearly right angles. The loops differ somewhat in conformation. Indeed, an X-ray structure of the glucocorticoid receptor domain bound to DNA shows that the configuration of cysteine ligands around the Zn has opposite 'hand' in the two elements²⁸. Nonetheless, the overall structure of the domain can probably still be thought of as the duplication of a loop-helix motif. The name 'double loop-Zn-helix' has been proposed for this structure²⁷, for which LZnH might be a suitable acronym. Residues near the N terminus of the first helix determine DNA specificity²⁹⁻³¹. Proteins containing these domains bind as dimers, and residues at the beginning of the second loop are involved in dimer contacts³⁰. The consensus DNA elements these proteins recognize have a corresponding two fold symmetry³⁰. In this respect, the class-2 Zn-binding domains resemble the prokaryotic HTH proteins. In the crystal structure²⁸, the glucocorticoid receptor binds with the first helix of its DNA-binding domain lying in the major groove, and side chains from the second and third turns of this helix interact with DNA bases. Positioning contacts to sugar-phosphate backbone on either side of the groove come from the initial loop and from the second loop-helix element. Dimer contacts between the two domains involve residues in the second loop.

All known examples of class-3 Zn domains are found in yeast transcriptional activators, of which GAL4 is the most studied³². GAL4 binds as a dimer³³ to a twofold-symmetric, 17-bp site^{34,35}. The DNA-binding domain is at the amino terminus: residues 1-74 are sufficient for recognition, and slightly larger fragments (such as 1-147) contain additional dimerization contacts that enhance specific affinity³³. There are six cysteines and two Zn ions per DNA-binding element. No three-dimensional structures have been completed, but NMR and X-ray crystallographic studies are in progress. NMR results show that the two Zn ions bound by a subunit form a binuclear cluster, with each Zn coordinated by four cysteines; two of the six cysteines are therefore shared¹⁹. The Zn ions are about 3.5 Å apart.

The so-called GATA-binding proteins, which include the haematopoietic-regulatory factor GATA-1, may contain a fourth class of Zn domain that recognizes DNA³⁶.

There are also two clearly identified classes of Zn domain that recognize RNA. One type is found in the retroviral *gag*

FIG. 1 Ribbon diagrams representing various DNA-binding domains. All the domains are drawn from the point of view of the DNA, which can be imagined to have its helix axis oriented roughly vertically. Amino and carboxy-termini are indicated by the letters N and C. The first four domains (a–d) have an HTH motif, which is highlighted in the drawing. a, The homeodomain^{3,4} is a single three-helix bundle. An initial α helix lies parallel to the first helix of the HTH motif. The second HTH helix is two turns longer than in the majority of prokaryotic examples. An NMR structure of the Antennapedia (*Antp*) homeodomain in solution shows a kink in this helix³, but there is no sign of such a perturbation in a crystal structure of the closely related *engrailed* homeodomain in complex with DNA⁴. It is not clear whether this kink represents a real difference between the structures. b, The *lac* repressor 'head-piece' is also a three-helix bundle, but the additional helix succeeds the HTH motif in the sequence⁵. c, The HTH in the repressor of phages λ ⁶ and 434 (ref. 7) is part of a 5-helix bundle: helices 1, 2 and 3 form a roughly triangular unit, and helix 4 closes off its hydrophobic interior. Helices 2 and 3 are the HTH. Helix 5, differently oriented in the λ and 434 proteins, serves as a dimerization element rather than as a major structural part of the HTH-containing domain. The 434 repressor domain (R1-69) is shown here. In λ repressor, an N-terminal arm serves as an additional recognition element. d, In the DNA-binding domain of CAP, there is a three-helix bundle closed off by a small β sheet⁸. If the HTH elements are superimposed, the first helix of the CAP bundle is spatially coincident with helix 4 in lambda and 434 repressors. The sheet is formed by the connection between the first helix and the HTH, together with a C-terminal hairpin. e, The Zn finger has one cysteine on each β strand, and the two histidines are on the inward-facing side of the helix. The Zn ion holds these two secondary structures against each other, and the conserved hydrophobic residues also contribute to stabilizing the domain. f, The glucocorticoid-receptor DNA-binding domain has two loop-helix modules, each with four cysteines coordinating a Zn^{2+} ion. The residues highlighted in pink are involved in dimer contacts when the domain binds to DNA (see Fig. 2d).



protein that binds viral genomic RNA. Its structure has been determined by NMR³⁷. It is unrelated to the DNA-binding structures just described. The other is in the *tat* protein of certain retroviruses such as human immunodeficiency virus³⁸. The Zn fingers of the class-1 protein TFIID also bind 5S RNA specifically³⁹. The structure of the 1:1 complex (known as the 7S particle) has not yet been determined, but analysis of mutational and protection data suggests that the mechanism of recognition is different from the TFIID/DNA interaction⁴⁰.

The leucine-zipper coiled-coil

A third type of DNA binding element was originally identified by its dimerization motif, an α -helical segment referred to as a 'leucine zipper'⁴¹. This ~30-residue segment forms the C-terminal half of a fragment that contains a positively charged 'basic region' at its N terminus. Because dimerization occurs through formation of an α -helical coiled-coil^{42,43}, I will refer to the complete structure as a 'bCC domain'. The name 'bZIP' is also used⁴⁴. Like all parallel α -helical coiled-coils, the leucine zipper has a heptad repeat with predominantly non-polar residues at the first and fourth positions. A leucine residue is found almost exclusively at the fourth position of four or five successive heptads in GCN4, Fos, Jun and other members of the bCC family⁴¹. The normal repeat of an α -helical coiled-coil is about 100 residues, so the α helices in the leucine zipper would be expected to wrap around each other by about one-third of a turn.

Properties of peptides representing the coiled-coil region alone, the basic region alone, and the entire bCC domain are

consistent with the following picture^{42–50}. Above an appropriate protein concentration, a dimer is formed by association of the leucine zipper segments. The unpaired chains do not form stable α helices, and the coil-to-helix transition is coupled to dimerization. The leucine zippers of some bCC proteins do not form stable homodimers, and a suitable partner (such as Jun in the case of Fos) is required for dimerization and DNA binding (see refs summarized in ref. 79). This property allows for more complex regulatory circuits than can be obtained with homodimers alone. In the absence of DNA, the basic region appears to be relatively unstructured, although at low temperatures some α -helix formation is detected⁴⁹. When bound to DNA, the basic region becomes almost entirely α -helical. Whether it forms a single helix or several helical segments is not yet known.

The bCC proteins recognize DNA sites 9–10 bp long^{51–53}. The consensus sequences are twofold symmetric—one set contains a central base pair (the Ap1 site: XTGACTCAX); the other set has its twofold symmetry between base pairs (the CREB/ATF site: XTGACGTCAX). Some bCC proteins, such as yeast GCN4, bind with nearly equal affinity to both Ap1 and CREB/ATF sites. Chemical protection and modification experiments indicate that contacts are made exclusively in the major groove over about 10 bp^{54–57}. The characteristics of the sequences just described imply that the protein must in some way wrap around DNA by about one half turn to either side of the centre of the site. Two related models for bCC binding have been

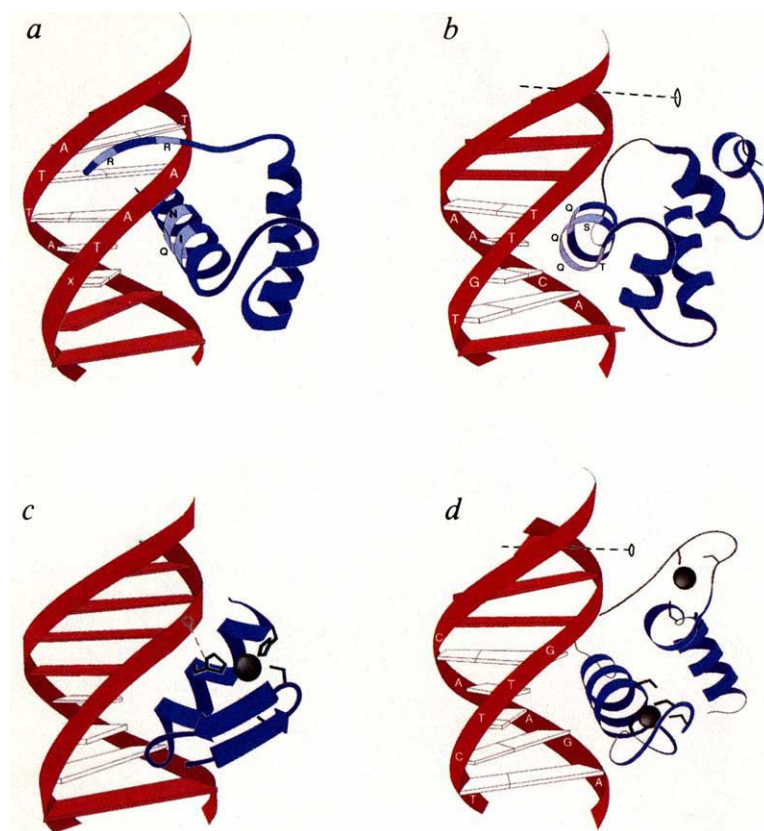


FIG. 2 DNA complexes of four of the domains illustrated in Fig. 1. The structures in *a*, *b* and *d* are viewed from a direction corresponding roughly to the right of the diagrams in Fig. 1: the Zn-finger complex (*c*) is drawn from a direction $\sim 180^\circ$ from the view in Fig. 1*e*. *a*, The engrailed homeodomain in complex with DNA containing a TAATX 'core consensus' sequence⁴. In the crystal structure, the sixth residue of the recognition helix (Ile) contacts the final T of this core sequence, and the tenth residue (Asn) contacts the penultimate A; minor-groove contacts to the first two core base pairs are made by residues in the N-terminal arm. All these residues, as well as the ninth recognition-helix residue (Gln), are highlighted. *b*, The 434 repressor DNA-binding domain (R1-69), bound to a half-site with the ACAA consensus sequence⁷². The twofold axis relating half sites is shown by a dashed line. Highlighted residues, critical for base-pair interactions, include the last residue in the turn (Thr), and residues 1, 2, 3 and 6 in helix 3 (Gln, Gln, Ser, Gln). Compare this diagram with the one in Fig. 2*a* to see that the position with respect to DNA of the HTH motif is quite different in the homeodomain and R1-69 complexes. *c*, A Zn-finger bound to DNA²². The N and C termini of the module project in opposition directions along the major groove. The first conserved His residue bridges between the Zn^{2+} and a DNA phosphate. The three base pairs that are contacted directly in the Zif 268 structure are highlighted. Note that the phosphate liganded by the His 'belongs' to the central base pair of the next three-base-pair set. *d*, The glucocorticoid receptor DNA binding domain recognizes a glucocorticoid response element (GRE) with consensus sequence AGAACAXXXTGTTCT (ref. 26). One half of a GRE is shown, together with a bound domain; the twofold axis relating the half-sites is shown as a dashed line. The thyroid hormone receptor recognizes a similar palindromic response element but with a shorter central spacer²⁷. The dimer contacts are likely to be quite different in that case.

proposed, in which α -helical basic regions diverge from the α -helical coiled-coil segment in order to fit into the major groove^{47,57}. At least one kink or break in the helix is probably needed to contact the base pairs most distal to the dyad, and available data cannot rule out another sharp kink at the basic region/leucine zipper junction. The absence of rigid structure in the basic regions of unbound bCC domains is important so that the protein can fit around the double helix.

β -ribbon proteins

There are two classes of prokaryotic proteins that interact with DNA through β -ribbon recognition elements⁵⁸. The first class includes the Met J repressor of *Escherichia coli* and the *arc* and *mnt* repressors of *Salmonella* phage P22. The second class includes the protein HU, which forms condensed nucleoprotein structures in many prokaryotes, the *E. coli* integration host factor IHF, and the *B. subtilis* phage transcription factor TF1. The latter two bind to specific sites.

The structure of Met J has been determined by X-ray crystallography⁵⁹, and the structure of Arc by NMR⁶⁰. These proteins are dimers with a core composed of four α helices, two from each subunit; met J contains an additional C-terminal helix (Fig. 3). An antiparallel β ribbon, formed from an N-terminal segment contributed by each monomer, protrudes from the core. In the specific DNA complex of met J that has been studied crystallographically, the β ribbon lies in the major groove⁵⁸. This recognition element requires dimerization for its folded structure. Naturally occurring operators contain at least two tandem repeats of an 8-bp 'met box'. The consensus met box sequence

is twofold symmetric, and one repressor dimer binds to each 8-bp element⁶¹. The crystal structure of the complex shows that the first of the core helices is so positioned that it contacts in antiparallel fashion a corresponding helix from the adjacent dimer⁵⁸. Binding is therefore cooperative.

HU and its relatives are believed to contact the minor groove of DNA. The structure of the HU has been determined crystallographically⁶². It has a core composed of two helices from each monomer, and two projecting β ribbons, each formed by the

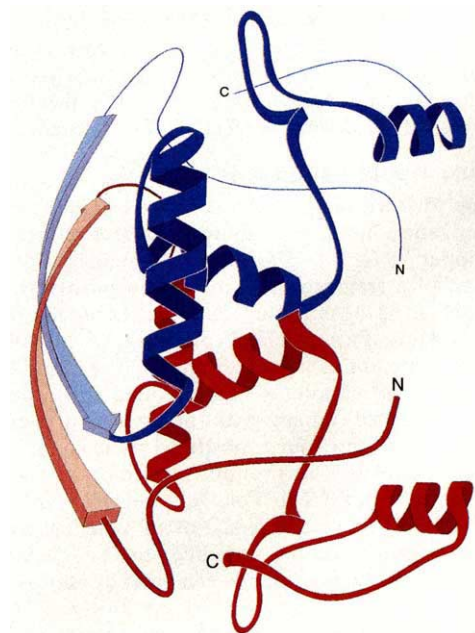


FIG. 3 The met J dimer⁵⁹. The β ribbon at the left (highlighted) lies in the major groove of DNA when the protein binds to a met box⁵⁸. The bound DNA would be oriented roughly vertically at the left of the diagram. The helix in the foreground interacts with a corresponding helix from another dimer, when the protein is bound cooperatively at adjacent sites.

C-terminal third of the polypeptide chain. The tips of these ribbons are disordered, presumably so that they can wrap around the DNA helix. Model building cannot distinguish between wrapping in the major or minor groove, but chemical protection studies show minor groove interactions for the homologous IHF (refs 63, 64).

The helix-loop-helix and other structures

The group of proteins with a segment known as a 'helix-loop-helix' have been characterized only by sequence regularities⁶⁵. No direct structural information is yet published. A basic region of about 15 residues lies immediately N-terminal to a 15-residue segment with conserved sequence characteristics typical of an amphipathic helix. A region of variable size (9–20 residues) separates this presumed helix from another of similar length. The helix-loop-helix element mediates dimerization. In almost all cases, a heterodimer of two such proteins is the active DNA-binding species.

There are a number of known DNA-binding proteins or protein domains that do not appear to fall into one of the groups described above. These include the heat-shock factor⁶⁶, various viral activators such as SV40 large T antigen⁶⁷ and papilloma E2 protein⁶⁸, NF- κ B (ref. 69), and TFIID (ref. 70). The taxonomy outlined here is therefore still incomplete.

Functional variety

Why do DNA-binding domains come in such a variety of structural types? Each type has some distinct functional properties. The DNA recognition element is a fixed (or relatively fixed),

preformed surface in many cases, but in others, it folds only when it binds. Thus, Zn fingers, steroid receptor domains, and prokaryotic HTH proteins have a 'recognition helix' that protrudes from a compact domain. By contrast, the bCC basic region and the IHF β ribbon have flexible segments that wrap around a DNA groove (major and minor, respectively), recognizing a set of base pairs immediately adjacent to a twofold axis. One module is generally not sufficient for site-specific recognition, because no one motif can contact more than a few base pairs. The oligomeric structure of a protein, or the covalent association of modules within a polypeptide chain, is therefore a critical characteristic. The prokaryotic HTH proteins, the steroid receptor-like class-2 Zn proteins, and the GAL4-like class-3 Zn proteins, form homodimers. The bCC (bZip) and HLH proteins seem to have evolved to form selective heterodimers. The class-1 Zn fingers form covalent concatemers. The length and symmetry of the DNA binding site correlates in an obvious way with quaternary organization of the protein. Further important factors in the evolution of binding-domain design may be the interaction with other proteins that bind at adjacent or intervening sites and (in the case of eukaryotic proteins) with the histone scaffolding of chromatin. Indeed, the structural requirements of such interactions, rather than the characteristics of the DNA site itself, may be particularly important for evolutionary selection of the different designs described here. □

Stephen C. Harrison is at the Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

- Pabo, C. O. & Sauer, R. T. *An. Rev. Biochem.* **53**, 293–321 (1984).
- Aggarwal, A. K. & Harrison, S. C. *A. Rev. Biochem.* **59**, 933–969 (1990).
- Qian, Y. Q. *et al. Cell* **59**, 573–580 (1989).
- Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. *Cell* **63**, 579–590 (1990).
- Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. *J. Mol. Biol.* **182**, 179–182 (1985).
- Jordan, S. R. & Pabo, C. O. *Science* **242**, 893–899 (1988).
- Mondragón, A., Subbiah, S., Almo, S. C., Drott, M. & Harrison, S. C. *J. molec. Biol.* **205**, 189–200 (1989).
- Weber, I. T. & Steitz, T. A. *J. molec. Biol.* **198**, 314–326 (1987).
- Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. *Nature* **317**, 782–786 (1985).
- Kostrewa, D. *et al. Nature* **349**, 178–180 (1991).
- Wharton, R., Brown, E. L. & Ptashne, M. *Cell* **38**, 361–369 (1984).
- Otting, G. *et al. EMBO J.* **9**, 3085–3092 (1990).
- Pabo, C. O. *et al. Science* **247**, 1210–1213 (1990).
- Kelleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53**, 927–936 (1988).
- Hanes, S. D. & Brent, R. *Cell* **57**, 1275–1283 (1989).
- Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
- Brown, R. S., Sander, C. & Argos, P. *FEBS Lett.* **186**, 271–274 (1985).
- Freedman, L. P. *et al. Nature* **334**, 543–546 (1988).
- Pan, T. & Coleman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2077–2081 (1990).
- Jacobs, G. & Michaels, G. *New Biologist* **2**, 583–584 (1990).
- Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 99–102 (1988).
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. *Science* **245**, 635–637 (1989).
- Klevit, R., Herriott, J. R. & Horvath, S. J. *Proteins* **7**, 215–226 (1990).
- Pavletich, N. P. & Pabo, C. O. *Science* **252**, 809–817 (1991).
- Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. *Nature* **349**, 175–178 (1991).
- Härd, T. *et al. Science* **249**, 157–160 (1990).
- Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. *Nature* **348**, 458–461 (1990).
- Luisi, B. *et al. Nature* **352**, 497–505 (1991).
- Umesono, K. & Evans, R. M. *Cell* **57**, 1139–1146 (1989).
- Mader, S., Kumar, V., deVerneuil, H. & Chambon, P. *Nature* **338**, 271–274 (1989).
- Danielson, M., Hinck, L. & Ringold, G. M. *Cell* **57**, 1131–1138 (1989).
- Johnston, M. *Microbiol. Rev.* **51**, 458–476 (1987).
- Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F. & Ptashne, M. *J. molec. Biol.* **209**, 423–432 (1989).
- Bram, R. & Kornberg, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 43–47 (1985).
- Ginger, E., Varnum, S. M. & Ptashne, M. *Cell* **40**, 767–774 (1985).
- Orkin, S. H. *Cell* **63**, 665–672 (1990).
- Summers, M. F., South, T. L., Kim, B. & Hare, D. R. *Biochemistry* **29**, 329–340 (1990).
- Frankel, A. D., Bredt, D. S. & Pabo, C. O. *Science* **249**, 70–73 (1988).
- Pelham, H. R. B. & Brown, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4170–4174 (1980).
- You, Q. & Ro maniuk, P. J. *Nucleic Acids Res.* **18**, 5055–5062 (1990).
- Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759–1764 (1988).
- O'Shea, E. K., Rutkowski, R. & Kim, P. S. *Science* **243**, 538–542 (1989).
- Oas, T. G., McIntosh, L. P., O'Shea, E., Dalquist, F. W. & Kim, P. S. *Biochemistry* **29**, 2891–2994 (1990).
- Shuman, J. D., Vinson, C. R. & McKnight, S. L. *Science* **249**, 771–774 (1990).
- O'Shea, E. K., Rutkowski, R., Stafford, W. F. & Kim, P. S. *Science* **245**, 646–648 (1989).
- Talanian, R. V., McKnight, C. J. & Kim, P. S. *Science* **249**, 769–771 (1990).
- O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. *Science* **249**, 775–778 (1990).
- Weiss, M. A. *et al. Nature* **347**, 575–578 (1990).
- Weiss, M. A. *Biochemistry* **29**, 8020–8024 (1990).
- Patel, L., Abate, C. & Curran, T. *Nature* **347**, 572–574 (1990).
- Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234**, 451–457 (1986).
- Lee, W., Haslinger, A., Karin, M., Tjian, R. *Nature* **325**, 368–372 (1987).
- Hai, T., Liu, F., Allegretto, E. A., Karin, M. & Green, M. *Genes Dev.* **2**, 1216–1226 (1988).
- Oakley, M. G. & Dervan, P. B. *Science* **248**, 847–850 (1990).
- Nye, J. A. & Graves, B. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3992–3996 (1990).
- Gartenberg, M. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6034–6038 (1990).
- Vinson, C. R., Sigler, P. B. & McKnight, S. L. *Science* **246**, 911–916 (1989).
- Phillips, S. E. V. *Curr. Opin. Struct. Biol.* **1**, 89–98 (1991).
- Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. *Nature* **341**, 705–710 (1989).
- Berg, J. N., van Ophemden, J. H. J., Burgering, M. J. M., Boelens, R. & Kaptein, R. *Nature* **346**, 586–589 (1990).
- Phillips, S. E. V. *et al. Nature* **341**, 711–714 (1989).
- Tanaka, I., Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. *Nature* **310**, 376–381 (1984).
- White, S. W., Appelt, K., Wilson, K. S. & Tanaka, I. *Proteins* **5**, 281–288 (1989).
- Yang, C.-C. & Nash, H. W. *Cell* **57**, 869–880 (1989).
- Murre, C., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777–783 (1989).
- Sorger, P. K. & Nelson, H. C. M. *Cell* **59**, 807–813 (1989).
- Arthur, A. K., Hoess, A. & Fanning, E. J. *Viral* **62**, 1999–2006 (1988).
- McBride, A. A., Schlegel, R. & Howley, P. M. *EMBO J.* **7**, 533–539 (1988).
- Kieran, M. *et al. Cell* **62**, 1007–1018 (1990).
- Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. *Cell* **58**, 1173–1181 (1989).
- Pabo, C. O. & Lewis, M. L. *Nature* **298**, 443–447 (1982).
- Aggarwal, A. K., Rodgers, D. W., Drott, M., Ptashne, M. & Harrison, S. *Science* **242**, 899–907 (1988).
- Mondragón, A., Wolberger, C. & Harrison, S. C. *J. molec. Biol.* **205**, 179–188 (1989).
- Mondragón, A. & Harrison, S. C. *J. molec. Biol.* **219**, 321–334 (1991).
- Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 744–749 (1981).
- Brennan, R. G., Roderick, S. L., Takeda, Y. & Matthews, B. W. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8165–8169 (1990).
- Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001–1007 (1991).
- Otwinowski, Z. *et al. Nature* **335**, 321–329 (1988).
- Kerppola, T. K. & Curran, T. *Current Opinions in Structural Biology* **1**, 71–79 (1991).

ACKNOWLEDGEMENTS. I thank R. Hellmiss for computer rendering of illustrations, C. Pabo, P. Wright, R. Kaptein and S. Phillips for sending coordinates, and T. Ellenberger, R. Marmorstein, C. Pabo, A. Johnson and M. Ptashne for critically reading the manuscript.

Forcing mechanisms of the Indian Ocean monsoon

Steven Clemens*, Warren Prell*, David Murray*,
Graham Shimmield[†] & Graham Weedon[‡]

* Geological Sciences, Brown University, Providence, Rhode Island 02912-1846, USA

[†] Edinburgh University, West Mains Road, Edinburgh EH9 3JW, UK

[‡] Earth Sciences, Cambridge University, Downing Street, Cambridge CB2 3EQ, UK

Sediments in the Arabian Sea provide biological, biogeochemical and lithogenic evidence of past changes in the Indian Ocean summer monsoon winds. For the past 350,000 years, this system has been externally forced by cyclical changes in solar radiation, and internally phase-locked to the transport of latent heat from the southern subtropical Indian Ocean to the Tibetan Plateau. In contrast to the results of general circulation models, these geological data suggest that the climate change associated with variability in global ice volume is not a primary factor in determining the strength and timing of the monsoon winds.

THE Indian Ocean summer monsoon represents one of the Earth's most dynamic interactions between atmosphere, oceans and continents. This interhemispheric phenomenon influences climate on a seasonal basis from eastern Africa through central Asia. The modern summer monsoon is driven by two primary mechanisms: differential (land-ocean) sensible heating and tropospheric latent heating. Differential sensible heating results in the seasonal formation of low atmospheric pressure over the Asian Plateau, high atmospheric pressure over the relatively cold southern subtropical Indian Ocean and, thus, the initiation of summer monsoon circulation (Fig. 1). Much of the intensity of the modern monsoon derives from direct heating of the troposphere through latent heat collected over the southern subtropical Indian Ocean, transported across the Equator and released by precipitation over the Asian Plateau¹. The combined effects of sensible and latent heating result in strong ($\sim 15 \text{ m s}^{-1}$) southwest monsoon winds over the Arabian Sea during June, July and August².

Summer monsoon winds over the western Arabian Sea result in the upwelling of cold, nutrient-rich waters accompanied by high productivity³⁻⁶, and in the aeolian transport of lithogenic material to the Arabian Sea from the surrounding deserts of Somali and Arabia⁷⁻⁹ (Fig. 1). Sediment-trap studies indicate that 70% of the biological and 80% of the lithogenic sediment in the northwestern Arabian Sea is deposited during the summer monsoon months¹⁰. Regional sediments therefore contain a geological record of variation in the strength and timing of summer monsoon circulation (Fig. 2).

Previous studies of summer monsoon variability (spanning the past 150,000 years) concentrated on foraminiferal^{11,12} and pollen^{12,13} evidence from the Arabian Sea in addition to comparisons between data and models using the National Center for Atmospheric Research Community Climate Model (NCAR CCM)^{14,15}. These studies suggest that stronger Arabian Sea monsoon winds in the past were associated with increased radiation in the Northern Hemisphere summer and reduced glacial-age boundary conditions (ice-sheet extent, continental albedo and sea surface temperature in the western Indian

Ocean). Sensitivity tests examining the relative influence of these two parameters indicate that the strength of the CCM monsoon winds in the Arabian Sea is twice as sensitive to changing glacial boundary conditions as to changes in solar radiation^{15,16}, whereas monsoon precipitation patterns over southern Asia are more sensitive to radiation changes¹⁵. Here we use biological, biogeochemical and lithological evidence from longer (350,000 years) sediment records to infer the physical relationships within the climate system that cause changes in monsoon wind strength, thus providing a more extensive framework for comparison with previous studies and with results from model simulations. Like previous CCM results, the geological records indicate that summer monsoon winds are externally forced by cyclical changes in the distribution of solar radiation. In contrast to previous CCM results, the geological data indicate that the timing of latent heat transport between the southern subtropical Indian Ocean and the Tibetan Plateau is a primary internal forcing mechanism and that only minor changes in monsoon winds are associated with variability in global ice volume.

Potential monsoon forcing mechanisms

Studies of monsoon dynamics reveal that three dominant mechanisms interact to determine the relative strength of the modern Indian Ocean monsoon. In addition to seasonal radiation patterns and the availability of latent heat (discussed

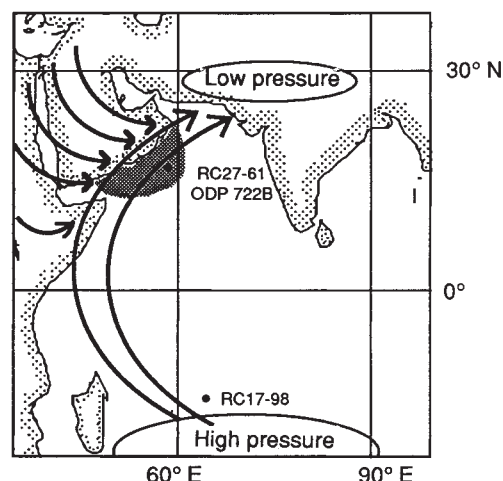


FIG. 1 Site location map and generalized atmospheric circulation during the summer monsoon months (June to August). Regional wind patterns (arrows) stem from the north-south pressure gradient and cyclonic circulation around the Asian low-pressure centre. Piston core RC27-61 ($16^{\circ} 38' \text{N}$, $59^{\circ} 52' \text{E}$, 1,890 m below sea level {m.b.s.l.}); Ocean Drilling Program site 722B ($16^{\circ} 37' \text{N}$, $59^{\circ} 48' \text{E}$, 2,028 m.b.s.l.); and piston core RC17-98 (13°S , $65^{\circ} 37' \text{E}$, 3,409 m.b.s.l.). Shaded area designates the region of highly productive coastal upwelling, driven by southwest monsoon winds blowing along the Arabian margin, and the area of maximal dust deposition from desert source areas in Somali and Arabia.

TABLE 1 Cross-spectral coherency and phase relationships with ETP

Record	Coherency (k) Phase angle ($^{\circ}$)*		
	100-kyr cycle (eccentricity; e)	41-kyr cycle (obliquity; ϵ)	23-kyr cycle (precession; $\Delta e \sin \omega$)
SPECMAP Stack	0.986 $-11^{\circ} \pm 5^{\circ}$	0.946 $-78^{\circ} \pm 11^{\circ}$	0.971 $-82^{\circ} \pm 8^{\circ}$
RC27-61	0.561	0.805	0.930
Grain size	$120^{\circ} \pm 39^{\circ}$	$-30^{\circ} \pm 22^{\circ}$	$-148^{\circ} \pm 12^{\circ}$
RC27-61	0.209	0.888	0.801
<i>G. bulloides</i> (%)†	$140^{\circ} \pm 69^{\circ}$	$4^{\circ} \pm 16^{\circ}$	$-121^{\circ} \pm 23^{\circ}$
ODP 722B	0.775	0.760	0.827
Opal flux	$116^{\circ} \pm 24^{\circ}$	$27^{\circ} \pm 25^{\circ}$	$-113^{\circ} \pm 21^{\circ}$
ODP 722B	0.852	0.676	0.954
Barium flux	$100^{\circ} \pm 19^{\circ}$	$-32^{\circ} \pm 31^{\circ}$	$-104^{\circ} \pm 10^{\circ}$
RC17-98	0.986	0.899	0.991
SST cold‡	$-173^{\circ} \pm 5^{\circ}$	$-15^{\circ} \pm 15^{\circ}$	$-123^{\circ} \pm 4^{\circ}$

Cross-spectral parameters: analyses, 6 to 350 kyr. 80% confidence interval for coherency is 0.787; 95% confidence interval for coherency is 0.913. Bandwidth, $BW = 0.007750$ (kyr^{-2}); interpolated time interval, $\Delta t = 2$ kyr.

* The phase error associated with our age models is $\sim \pm 2$ kyr ($\pm 7^{\circ}$ for 100-kyr cycle, $\pm 18^{\circ}$ for 41-kyr cycle, $\pm 31^{\circ}$ for 23-kyr cycle). Positive values indicate phase leads with respect to orbital maxima.

† Linear trend removed.

‡ Estimates of maximum cold temperatures for Northern Hemisphere summer (Southern Hemisphere winter); excludes samples from 288 to 302 kyr to avoid possible effects of dissolution on SST estimates.

above), anomalously heavy snow cover in Asia has been linked to delayed-onset and weakened summer monsoons through the effects of increased continental albedo (decreased sensible heating)^{11,17-20}. Of these three potential forcing mechanisms, only seasonal radiation can be directly calculated for the past

350,000 years; the other two are inferred from the geological record.

Solar radiation patterns, for any given latitude and season, can be calculated from long-term variations in the eccentricity (e) of the Earth's orbit, the tilt of the Earth's axis with respect to the plane of the ecliptic (obliquity; ϵ), and the precession index ($\Delta e \sin \omega$; where ω is the longitude of perihelion measured from the moving vernal point)^{21,22}. A frequency spectrum of a normalized, summed record of eccentricity, obliquity (tilt) and precession (ETP) illustrates that these three orbital parameters have characteristic periods of 100,000, 41,000 and 23,000 years respectively (Fig. 3a)²³. Virtually all the variance in the distribution of solar radiation, however, is associated with the obliquity and precession parameters alone^{24,25}. Thus, over orbital time scales, periods of increased Northern Hemisphere summer radiation, associated with increased precession and obliquity, should result in strengthened summer monsoon circulation as a function of increased differential sensible heating.

The marine $\delta^{18}\text{O}$ record provides a proxy for large-scale glacial-age conditions including continental snow and ice volume²⁶, and continental albedo (Fig. 3b)^{27,28}. In this respect, we use the SPECMAP $\delta^{18}\text{O}$ record²⁸ as a proxy for the effectiveness of sensible heating over the Asian Plateau with possible effects on monsoon circulation. Assuming changes in global ice volume are representative of changes in the vicinity of the Himalayas and Tibetan Plateau, then maximum global ice corresponds to maximum snow and/or ice cover and, therefore, to maximum albedo. Increased albedo reduces the effectiveness of sensible heating, resulting in weakened monsoon circulation^{11,14,29}. If monsoon winds are sensitive to environmental conditions associated with global ice-volume changes, then conditions of maximum ice should correspond to minimum monsoon strength (and vice versa), and monsoon tracers from the geological record should possess the 100,000 year variability known to dominate marine $\delta^{18}\text{O}$ records²⁴.

We employ sea-surface-temperature (SST) estimates³⁰, based on foraminiferal census counts from piston core RC17-98, as a proxy for latent heat availability (Fig. 3c). Recent studies point to the southern subtropical Indian Ocean as the source for 70% of the latent heat released over the Asian Plateau during the summer monsoon³¹⁻³³. Meteorological studies indicate that, over annual and interannual timescales, colder southern subtropical SSTs during boreal summer (austral winter) are associated with increased evaporation (latent heat flux from ocean to atmosphere), increased Southern Hemisphere trade winds, increased cross-equatorial transport of latent heat and strong summer monsoon winds over the Arabian Sea^{34,35}. The association between cold SSTs and increased evaporation can be attributed to large temperature and humidity gradients created by cold, dry winds blowing over relatively warm ocean waters. These gradients serve to decrease SST and at the same time collect latent heat from the ocean surface. Thus, colder SSTs at RC17-98 are interpreted as indicating greater availability of latent heat, and should be associated with strengthened monsoon circulation⁹. We next compare orbital radiation patterns, global ice-volume and southern subtropical SST records with four geological records of Arabian Sea monsoon wind strength in an effort to examine their relative influences on the strength and timing of monsoons in the late Quaternary period.

Monsoon tracers from the geological record

We examine records of biogenic opal flux, the concentration of the planktonic foraminifer *Globigerina bulloides* (relative to 33 other species), excess barium flux and lithogenic grain size; these four tracers are linked to monsoon strength through independent pathways in the marine and atmospheric systems (Fig. 2). Similar responses from independent monsoon tracers enable us to evaluate more confidently the temporal relationships between past monsoon strength and changes in both

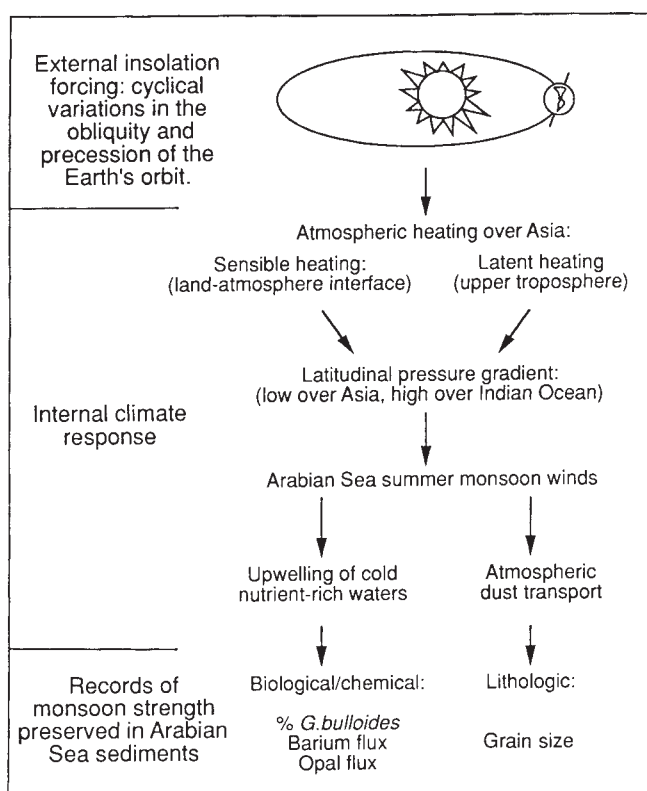


FIG. 2 Conceptual diagram linking the flow of solar radiation through the monsoon climate system with monsoon tracers preserved in the deep-sea sediment record.

external (solar) and possible internal forcing mechanisms (such as latent heating and glacial boundary conditions).

The monsoon tracers presented in Fig. 4 were measured from deep-sea sediment cores RC27-61 and ODP 722B which were recovered from the crest of the Owen Ridge, ~300 km off the coast of the Arabian Peninsula (Fig. 1). Chronostratigraphies for these records were developed by correlation of oxygen isotope events to the SPECMAP stacked oxygen isotope record²⁸ and are given in detail in Clemens and Prell^{9,36}. The average temporal resolution of these records is ~5,000 years.

The planktonic foraminifer *G. bulloides* is typically a subpolar species, but is also found in high abundance in tropical upwelling regions characterized by cold, nutrient-rich water (Fig. 4a)^{11,29,37–39}. High opal flux (radiolaria and diatoms) is associated with high productivity in much of the modern ocean^{40–42} and is thus linked to monsoon circulation as an index of silica productivity associated with monsoon-induced upwelling (Fig. 4b)^{10,43}. The grain size of the lithogenic component is linked to monsoon wind strength through the transport capacity⁷ of the low-level southwest monsoon and Shamal wind systems (Figs 1, 4c). Stronger winds transport larger lithogenic dust particles from desert source areas of the Somali and Arabian peninsulas⁹. Barium, concentrated on the tests of marine plankton (primarily diatoms), is found in elevated levels in sediments underlying highly productive waters^{44,45}. This biogeochemical tracer is useful because of its resistance to dissolution processes that affect both carbonate and siliceous sedimentary components. Excess barium flux, normalized for the terrestrial barium in aeolian

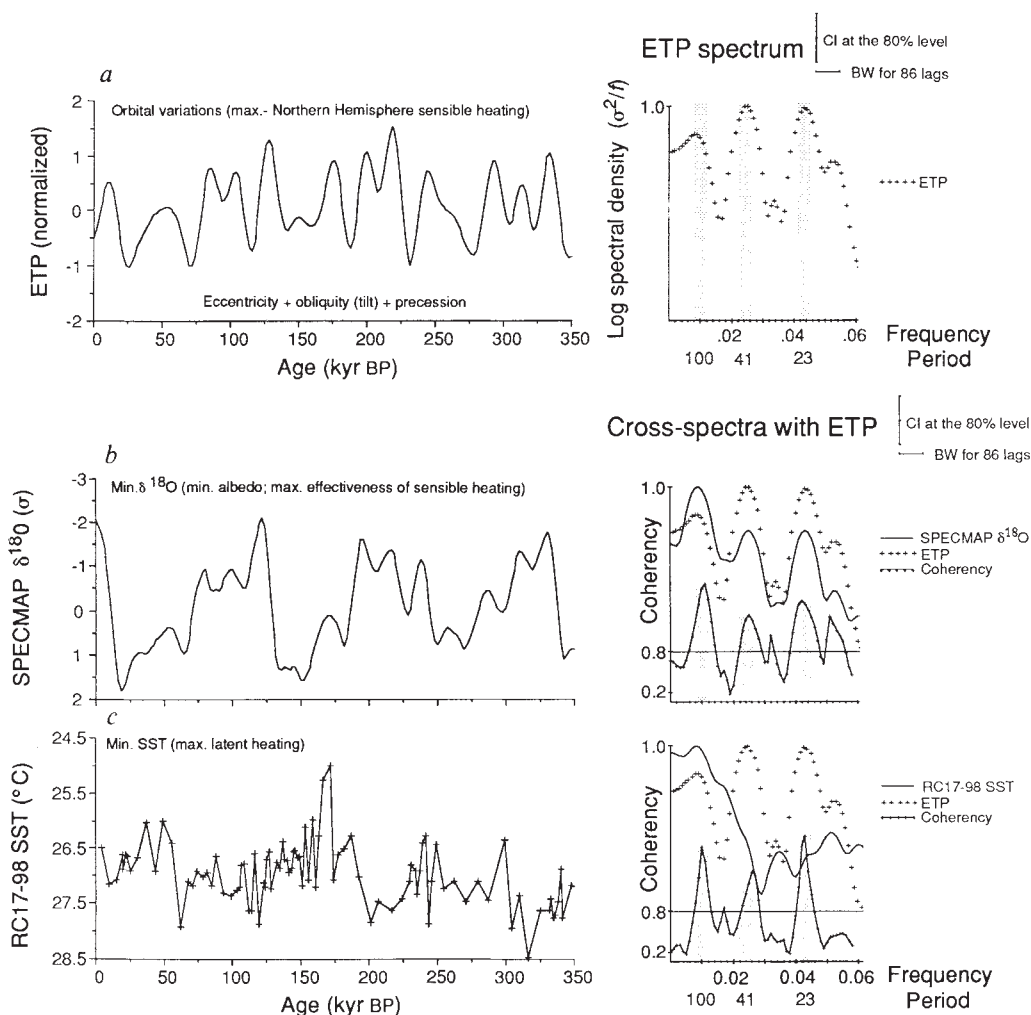
clays, provides a proxy record of marine biogenic barium production in the Arabian Sea (Fig. 4d)^{10,46}.

Although the overall character of these records is different, a visual comparison of the four monsoon tracers indicates simultaneous large-scale responses related to upwelling-induced productivity and monsoon wind strength. For example, relative amplitude differences notwithstanding, strong events are indicated in most records at (roughly) 9, 50, 123, 215, 261 and 329 kyr BP (Fig. 4). Minor differences in the exact timing and relative amplitudes of such events can be attributed to mechanisms not necessarily directly related to monsoon variation including differential opal and carbonate dissolution, dust source-area changes, sediment winnowing and ecosystem dynamics. Such mechanisms may also account for the more obvious differences between the records including, for example, the long-term increase toward the present in *G. bulloides*, the lack of a strong 9-kyr response in the grain size, and the intervals of damped amplitude in the opal and barium records (for example 60–100 kyr BP and 125–200 kyr BP). These differences will ultimately allow examination of the secondary mechanisms that drive the response of each individual tracer. Here, however, we concentrate on the first-order response to monsoon forcing. Although visual inspection is useful, frequency-domain analyses of these complex time series can provide additional insight into the dynamics of the monsoon system.

Frequency domain comparisons

Statistical comparisons of the three potential forcing mechanisms (Fig. 3) and the four monsoon tracers (Fig. 4) are made

FIG. 3 Three potential monsoon forcing mechanisms. **a**, External radiation. Variations in past patterns of seasonal radiation are summarized by a normalized, summed time series of the Earth's eccentricity, obliquity (tilt) and precessional parameters (ETP). Because increased radiation during boreal summer is associated with stronger monsoons, we reverse the sign of precession and reference it to Northern Hemisphere summer (June). Possible internal mechanisms for forcing the monsoon include **(b)** changes in the albedo of the Asian Plateau and thus changes in the effectiveness of sensible heating (as monitored by the marine $\delta^{18}\text{O}$ record of global ice volume) and **(c)** the availability of latent heat (as monitored by estimates of sea surface temperature from the southern subtropical Indian Ocean; colder SSTs indicate increased availability of latent heat). Cross-spectral comparisons with ETP indicate that SPECMAP $\delta^{18}\text{O}$ and RC17-98 SST are both coherent with changes in orbital radiation patterns. Spectral densities (σ^2/f , where σ^2 =variance and frequency $f=1/\text{period}$) are normalized and plotted on logarithm scales. The coherence spectrum (y-axis; solid line with crosses) is plotted on a hyperbolic arctangent scale. The solid horizontal line indicates confidence at the 80% level. Shaded bars highlight statistically significant portions of the coherence spectrum that occur over the primary orbital bands. Spectral parameters, coherence estimates and phase estimates are reported in Table 1.



in the frequency domain using cross-spectral analysis^{47,48}. Spectral peaks indicate the primary periods (1/frequency) at which variance is concentrated. Cross-spectral results are discussed in terms of coherency and phase. Coherency (k) is a measure of the linear correlation between two time series over a given frequency band when the phase difference is set to zero. Statistically significant coherence over a given frequency band indicates a strong linear relationship between two variables. Phase estimates over a specific frequency band indicate temporal (lead/lag) relationships between variables. Coherence estimates, in conjunction with the phase relationships, thus allow one to draw inferences about physical relationships between potential forcing mechanisms and climate response.

Cross-spectral analyses indicate that both the $\delta^{18}\text{O}$ record (albedo effect) and the southern subtropical SST record (latent heating) are coherent with concentrations of orbital variance at the primary orbital periods of 100,000, 41,000 and 23,000 years (Fig. 3). In addition to variance at the primary orbital periods, the SST record contains variance at heterodyne frequencies of the primary orbital periods (for example at 35,000 and 54,000 years) resulting in ill-defined primary peaks¹⁶. But the stability of the phase spectrum (not shown) indicates that the coherency between the SST and orbital parameters is not an artefact of

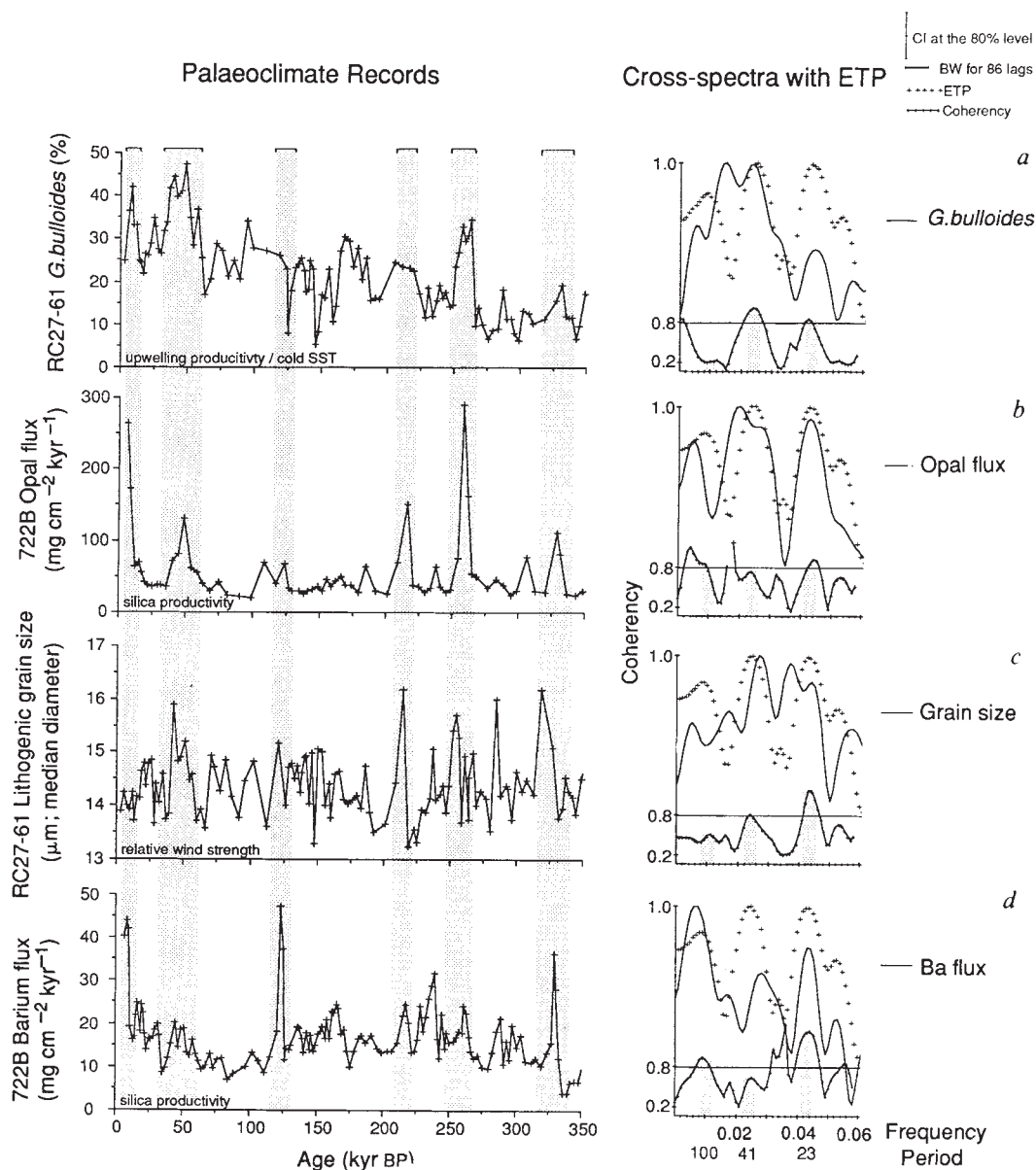
the computation, as might be inferred from the lack of well-defined peaks in the SST spectrum.

The four monsoon tracers demonstrate varying degrees of coherence with orbital parameters (Fig. 4). Strong coherencies are recorded over the precession band, moderately strong coherencies are recorded over the obliquity band, and moderate to weak coherencies are recorded over the eccentricity band. *G. bulloides* and grain size show negligible coherence with eccentricity, whereas opal and barium record moderate coherencies, although the opal coherence is associated with poorly aligned spectral peaks.

The spectral coherencies presented above demonstrate that the four monsoon tracers as well as the two potential internal forcing mechanisms are linearly related to external radiation forcing over the obliquity and precession bands. To estimate the relative influences of external and internal forcing mechanisms on monsoon variability, however, one must also consider coherency and phase relationships between the individual monsoon tracers and each potential internal forcing mechanism (a total of 14 cross-spectra). We summarize and illustrate all cross-spectral results (coherence and phase relationships) on phase wheel plots (Fig. 5).

The precessional phase wheel (Fig. 5a) illustrates strong

FIG. 4 Records of monsoon strength preserved in Arabian Sea sediments. These four monsoon tracers indicate similar first-order responses to monsoon forcing as illustrated by the similar timing of events at roughly 9, 50, 123, 215, 261 and 329 kyr BP (shaded bars). Stronger monsoons result in increases in (a) the percentage of *G. bulloides*, (b) opal flux, (c) lithogenic grain size and (d) barium flux. Cross-spectral comparisons with ETP indicate strong coherencies with precession, moderate coherencies with obliquity and moderate-to-weak coherencies with eccentricity. See Fig. 3 for explanation of cross-spectral plots. Spectral parameters, coherence estimates and phase estimates are reported in Table 1.



coherence between all four monsoon tracers and variance associated with precessional radiation. Although minor lead/lag relationships exist between the monsoon tracers (probably accounted for by processes not directly related to monsoon variability), these differences are within the error of the phase estimate and are thus indistinguishable (Table 1). The average phase is taken as the best estimate of the monsoon response relative to radiation maxima. The average of the four monsoon tracers indicates that maxima in monsoon strength occur $\sim 8,000$ years (-122°) after maxima in precessional radiation (an 8,000-year phase lag; $122^\circ/360^\circ \times 23,000$ years). This lag confirms that mechanisms internal to the climate system set the phase of strong monsoons over the precessional frequency band^{9,11}; if radiation were the sole forcing one would expect a near-zero or a slight phase lag⁴⁹. In addition to coherence with orbital precession, all four monsoon tracers demonstrate statistically significant coherence ($\geq 80\%$ confidence interval) with both SST minima (maximum latent heat availability) and ice-volume minima (minimum albedo; maximum effectiveness of sensible heating) as illustrated by dashed vectors on the precessional phase wheel (Fig. 5a). Maxima of monsoon strength occur $\sim 2,500$ years after ice-volume minima and simultaneously with minimum cold-season SSTs in the southern subtropical Indian Ocean. Thus, the 8,000-year phase lag between radiation maxima and monsoon maxima can be accounted for, in part, by increased

effectiveness of sensible heating, but, as a whole, by increased latent heating. This is consistent with modern annual and inter-annual coupling between stronger monsoons and increased ocean-atmosphere latent heat flux from the southern subtropical Indian Ocean^{34,35}. Southern subtropical SST minima lead precessional radiation minima in the southern subtropics (boreal summer/austral winter) by $\sim 57^\circ$ and are thus consistent with orbital radiation changes, but cannot be entirely accounted for by them, (Fig. 5a). Additional mechanisms must have a role in determining the SST phase.

Over the obliquity band, strong monsoons occur simultaneously with maxima in Northern Hemisphere radiation and minima in the southern subtropical SST (maximum latent heat) although coherencies are slightly weaker than those over the precession band (Fig. 5b). Within the error, the phases of all monsoon tracers are indistinguishable, with the exception of opal flux which reaches a maximum slightly earlier than the other three. The average monsoon response lags obliquity-induced radiation maxima by 8° , a statistically insignificant amount. This near-zero phase indicates a positive, linear relationship between increased radiation (sensible heating) and increased monsoon strength. The near-zero phase relationship between southern subtropical SST minima and maxima in the monsoon strength indicates that latent heating may also play a part in setting the phase of strong monsoons over the obliquity

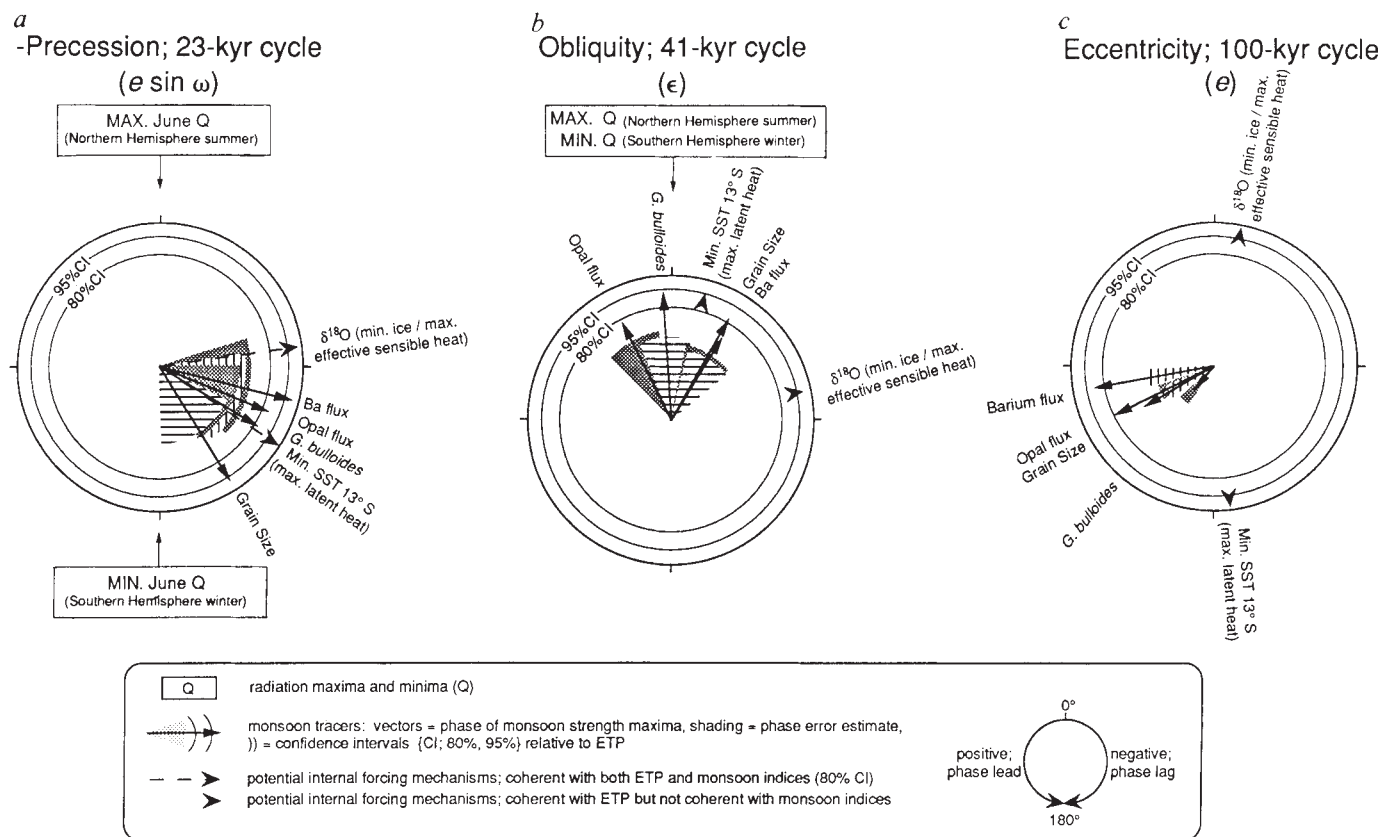


FIG. 5 Coherence and phase summary. The purpose of the phase wheel diagram is to display all the available cross-spectral relationships between orbital variations and changes in climate tracers that occur over a given frequency band. Zero phase on each wheel is set at (a) maximum precession (referenced to June perihelion; reference to July/August would rotate all phases $30^\circ/60^\circ$ clockwise but would not alter the conclusions of this work), (b) maximum obliquity and (c) maximum eccentricity. Vectors are used to represent maxima in climate responses relative to orbital maxima, with positive angular distances from zero corresponding to the measured phase difference (+/counterclockwise) represents a lead; -(clockwise) represents a lag. The lengths of phase vectors indicate the degree of coherence with a given orbital parameter. Significance levels, at the 80% and 95% confidence

intervals (CI), are given by the inner and middle circles. Dashed arrows indicate significant coherence between monsoon tracers and potential internal forcing mechanisms ($\geq 80\%$ CI). For example, over the precession band, $\delta^{18}O$ is coherent with precession above the 95% CI and with all four monsoon tracers (above the 80% CI). Ice-volume minima occur $\sim 5,200$ yr (-82°) after maxima in precessional radiation ($82^\circ/360^\circ \times 23,000$ yr = 5,200 yr). Coherence and phase estimates with respect to ETP are reported in Table 1. The phases of relevant radiation maxima and minima (Q) are denoted in boxes. Phase errors between cores are $\sim \pm 2,000$ yr ($\pm 31^\circ$ for precession, $\pm 18^\circ$ for obliquity, $\pm 7^\circ$ for eccentricity) and are represented as shaded areas centred on phase wheel vectors.

band. In this case, obliquity maxima can account for both sensible heating maxima in the Northern Hemisphere and SST minima in the southern tropics⁵⁰. We point out, however, that coherencies between the southern subtropical SST record and the monsoon tracers fall below the 80% confidence interval over this band (SST vector not dashed; Fig. 5b). This relatively weak statistical link may be attributed to the geographical location of the southern subtropical SST record (13° S). We speculate, from the zero-phase relationship, that SST estimates from higher southern latitudes, under greater influence of obliquity forcing, may show significantly greater coherence with the monsoon tracers. Previous work on the lithogenic record¹⁶ evaluated the importance of heterodynes of the primary orbital elements occurring with periods of 35,000 and 54,000 years. Similar concentrations of variance near the 41,000-year spectral peak occur in the biological and geochemical records (Fig. 5a, b, d). The relatively weak cross-spectral coherence between the monsoon tracers and obliquity radiation may be due to interference with these heterodyne peaks. Over the obliquity band, we rule out change in environmental conditions associated with variability in global ice volume as a primary forcing mechanism because of the low coherence with the monsoon tracers ($\delta^{18}\text{O}$ vector not dashed; Fig. 5c), and because ice-volume minima (minimum albedo, maximum effectiveness of sensible heating) occur ~8,000 years (70°) after monsoon maxima, an impossible forcing-response relationship.

With the exception of barium flux, the monsoon tracers show relatively little variability at the 100,000 year period (Figs 4a–d, 5c). The 100,000 year variance in productivity indicated by the barium flux record suggests a control on the nutrient supply which may not necessarily be related to monsoon intensity⁵¹. Overall, however, the monsoon tracers show relatively low and variable coherence with orbital eccentricity in addition to low coherence with both the $\delta^{18}\text{O}$ and the southern subtropical SST records ($\delta^{18}\text{O}$ and SST vectors not dashed; Fig. 5c). A visual comparison of the monsoon tracers with the ice-volume record (compare Fig. 3b with 4a–d) indicates that this weak coherence can be, in part, attributed to very strong monsoons at ~50 and

~261 kyr BP, despite the relatively large extant ice-volume at these times. In addition, ice-volume minima (maximum effectiveness of sensible heating) lag monsoon maxima by ~33,000 years (119°) over the eccentricity band, an impossible forcing-response relationship (Fig. 5c). Given the fast-response physics linking evaporation and monsoon strength, there must also be a near-zero phase relationship if latent heating is to be considered a primary influence. Clearly, this is not the case, as southern subtropical SST minima significantly lead maxima in Arabian Sea monsoon strength (Fig. 5c). From these relationships, we infer that the small amount of variance concentrated at the 100,000 year band cannot be accounted for by environmental conditions associated with global ice-volume fluctuations, nor by the availability of latent heat.

Conclusions

Biological, biogeochemical and lithogenic evidence from the Arabian Sea sediment record indicates that Arabian Sea summer monsoon winds respond in a strong, coherent fashion to changes in the distribution of solar radiation as determined by cyclical variability in the obliquity and precession of the Earth's orbit. This is consistent with results from CCM simulations. The monsoon records demonstrate uniformly strong coherence with precession and moderate coherence with obliquity. Phase relationships, however, demonstrate that radiative forcing alone cannot account for the timing of strong monsoons over the past 350,000 years. Similar to modern annual and interannual monsoon dynamics, our data suggest that monsoon maxima over the precessional (and possibly the obliquity) bands are phase-locked to the cross-equatorial transport and release of latent heat over the Asian Plateau. Phase relationships and weak coherence between records of monsoon strength and global ice volume suggest that the influence of large-scale climate change associated with glacial–interglacial variability is not a primary factor in determining the relative strength and timing of Arabian Sea summer monsoon winds. This result is inconsistent with CCM results which suggest a strong sensitivity to glacial–interglacial variability. □

Received 6 June; accepted 10 September 1991.

- Webster, P. J. *The Elementary Monsoon* (Wiley, New York, 1987).
- Hastenrath, S. & Lamb, P. J. *Climatic Atlas of the Indian Ocean* (University of Wisconsin Press, Madison, 1979).
- Currie, R. I., Fisher, A. E. & Hargreaves, P. M. in *The Biology of the Indian Ocean* (eds Zeitzechel, B.) 37–52 (Springer, New York, 1973).
- Smith, R. L. & Bottero, J. S. *On Upwelling in the Arabian Sea* (Pergamon, New York, 1977).
- Prell, W. L. & Curry, W. B. *Oceanol. Acta* **4**, 91–98 (1981).
- Prell, W. L., Marvill, R. E. & Luther, M. E. *Paleoceanography* **5**, 447–457 (1990).
- Tsoar, H. & Pye, K. *Sedimentology* **34**, 139–153 (1987).
- Sirocko, F. & Sarnthein, M. in *Palaeoclimatology and Palaeometeorology: Modern and Past Patterns of Global Atmospheric Transport* (eds Leinen, M. & Sarnthein, M.) 401–434 (Kluwer, Boston, 1989).
- Clemens, S. & Prell, W. L. *Paleoceanography* **5**, 109–145 (1990).
- Nair, R. R. et al. *Nature* **338**, 749–751 (1989).
- Prell, W. L. in *Milankovitch and Climate* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 349–366 (Reidel, Hingham, 1984).
- Prell, W. L. & Van Campo, E. *Nature* **323**, 526–528 (1986).
- van Campo, E., Duplessy, J. C. & Rossignol-Strick, M. *Nature* **296**, 56–59 (1982).
- Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
- Prell, W. L. & Kutzbach, J. E. *J. geophys. Res.* **92**, 8411–8425 (1987).
- Clemens, S. C. & Prell, W. L. *J. geophys. Res.* (in press).
- Hahn, D. & Shukla, J. *J. Atmos. Sci.* **33**, 2461–2463 (1976).
- Yeh, T. C., Wetherald, R. T. & Manabe, S. *Mon. Weath. Rev.* **111**, 1013–1024 (1983).
- Dickson, R. R. *J. Clim. appl. Meteor.* **23**, 171–173 (1984).
- Shukla, J. in *Monsoons* (eds Fein, J. & Stephens, P.) 399–464 (Wiley, New York, 1987).
- Berger, A. L. *Quat. Res.* **9**, 139–167 (1978).
- Shackleton, N. J., Berger, A. & Peltier, W. R. *Trans. R. Soc. Edinburgh: Earth Sci.* **81**, 251–261 (1990).
- Imbrie, J. *J. geol. Soc. Lond.* **142**, 417–432 (1985).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Berger, A. L. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Shackleton, N. J. *Nature* **215**, 15–17 (1967).
- CLIMAP. *Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum* (Geological Society of America, Boulder, Colorado, 1981).
- Imbrie, J. et al. in *Milankovitch and Climate, Part 1* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Reidel, Hingham, 1984).
- Prell, W. L. in *Climate Processes and Climate Sensitivity* (eds Hansen, J. & Takahashi, T.) 48–57 (American Geophysical Union, 1984).
- Prell, W. L. *The Stability of Low-Latitude Sea-Surface Temperatures: an Evaluation of the CLIMAP Reconstruction with Emphasis on the Positive SST Anomalies* (U.S. Department of Energy, Washington, DC, 1985).
- Cadet, D. L. & Reverdin, G. *Mon. Weath. Rev.* **109**, 148–158 (1981).
- Cadet, D. & Reverdin, G. *Tellus* **33**, 476–487 (1981).
- Hastenrath, S. *Climate and Circulation of the Tropics* (Reidel, Boston, 1985).
- Cadet, D. L. *Mon. Weath. Rev.* **111**, 95–108 (1983).
- Cadet, D. L. & Diehl, B. C. *Mon. Weath. Rev.* **112**, 1921–1935 (1984).
- Clemens, S. C. & Prell, W. L. *Proc. ODP Sci. Res.* **117**, 365–388 (1991).
- Hutson, W. H. & Prell, W. L. *J. Paleontol.* **54**, 381–399 (1980).
- Cullen, J. L. & Prell, W. L. *Mar. Micropaleontol.* **9**, 1–52 (1984).
- Peterson, L. C., Overpeck, J. T., Kipp, N. G. & Imbrie, J. *Paleoceanography* **6**, 99–119 (1991).
- DeMaster, D. J. *Geochim. cosmochim. Acta* **45**, 1715–1732 (1981).
- Heath, R. G. *Dissolved Silica in Deep-Sea Sediments*, 77–93 (Society of Economic Palaeontologists and Mineralogists, Tulsa, 1974).
- Leinen, M. et al. *Geology* **14**, 199–203 (1986).
- Murray, D. W. & Prell, W. L. *Proc. ODP Sci. Res.* **117**, 343–364 (1991).
- Dymond, J. *Geol. Soc. Am. Mem.* **154**, 133–173 (1981).
- Bishop, J. K. B. *Nature* **332**, 341–343 (1988).
- Shimmield, G. B. & Mowbray, S. R. *Proc. ODP Sci. Res.* **117**, 409–429 (1991).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden Day, Oakland, 1968).
- Imbrie, J., McIntyre, A. & Mix, A. in *Climate and Geosciences* (eds Berger, A., Schneider, S. & Duplessy, J.-C.) 121–164 (Kluwer, Boston, 1989).
- Short, D. A. & Mengle, J. G. *Nature* **323**, 48–50 (1986).
- Short, D. A., Mengle, J. G., Crowley, T. J., Hyde, W. T. & North, G. R. *Quat. Res.* (1991).
- Weedon, G. P. & Shimmield, G. B. *Proc. ODP Sci. Res.* **117**, 431–444 (1991).

ACKNOWLEDGEMENTS. We thank J. Kutzbach for a review. This work was supported by NSF and the Joint Oceanographic Institutions.

Peptide-binding specificity of the molecular chaperone BiP

Gregory C. Flynn, Jan Pohl*, Mark T. Flocco[†] & James E. Rothman

Program in Cellular Biochemistry and Biophysics, Rockefeller Research Laboratory, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10021, USA

* Microchemical Facility, Emory University School of Medicine, Atlanta, Georgia 30322, USA

[†] Microchemistry Core Facility, Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

Members of the heat-shock protein family (hsp70s) can distinguish folded from unfolded proteins. This property is crucial to the role of hsp70s as molecular chaperones and is attributable to the amino-acid specificity of the peptide-binding site. The specificity for peptide ligands is investigated using a set of peptides of random sequence but defined chain length. The peptide-binding site selects for aliphatic residues and accommodates them in an environment energetically equivalent to the interior of a folded protein.

How do cells distinguish folded from unfolded proteins? Proteins that fold in the endoplasmic reticulum are efficiently secreted; proteins that fail to fold remain behind and are rapidly degraded¹. A similar editing system operates in the cytoplasm to remove misfolded or denatured proteins². Newly synthesized proteins, and proteins freshly translocated across membranes, are recognized and engaged by a folding machinery that prevents aggregation reactions; already folded proteins are apparently ignored^{3,4}. The biophysical basis for these kinds of distinctions is of fundamental significance for both cell biology and protein structure, as it must embody the cell's own definition of a folded state⁵.

This specificity is exhibited by the polypeptide-binding site of members of the hsp70 (heat-shock protein of relative molecular mass 70,000 (70K)) family of molecular chaperones. There is much evidence to suggest that hsp70s are a requisite part of the cell's protein folding, degradation and, possibly, repair machinery^{4,6-10}. These binding proteins coat polypeptide chains as they emerge from ribosomes or membranes, preventing aggregation reactions and premature folding. Palleros *et al.*¹¹ confirmed that an hsp70 can recognize only unfolded forms of a protein. Here we directly measure the amino-acid side-chain specificity of this binding site, and find that it has virtually the same binding energy 'fingerprint' as the interior of a folded protein, providing a simple biophysical basis for the specificity of the recognition.

Peptide-binding affinity

We reported earlier that hsp70 family members can bind short peptides and that bound peptide is released on hydrolysing a molecule of ATP¹². This parallels the ability of hsp70s to associate in cells with misfolded proteins, and to dissociate from them in the presence of ATP *in vitro*. Bound peptides triggered cycles of binding and release that resulted in a peptide-dependent ATPase activity. The K_m s of the peptides varied over a range of 10 μ M to >1 mM, but no clear pattern of specificity was indicated.

Here we use an improved stochastic method to investigate the binding specificity of the hsp70, BiP. A series of peptides of known chain length but undefined sequence were prepared using a mixture of derivatized amino acids at each step of the

coupling in solid-phase synthesis; this produces a nested set of random oligomeric peptides of increasing chain length. The first step is to measure binding affinity as a function of chain length to determine the chain length that can be accommodated most efficiently in the peptide-binding site; longer chains will bind with reduced affinity. The use of random oligomers of defined length effectively eliminates sequence as a variable by averaging and leaves chain length as the only determinant of binding. The second step is to isolate complexes of the binding protein with the shortest random peptide that fills the binding pocket (to maximize information on any positional specificity) and to sequence the mixture of bound peptides. The degree of enrichment (or exclusion) of each amino acid from a given position in the bound oligomer compared with the input mixture then provides a simple statistical profile of the side-chain specificity of each position in the binding pocket. The relative abundance of each amino acid in the bound versus free peptide population provides a quantitative energy 'fingerprint' of the binding site, as the energy needed to transfer an amino acid into a given site in the binding pocket can be calculated.

To determine the chain length-dependence of the peptide-binding site of BiP, we prepared the peptides by solid-phase synthesis using the Fmoc-amino acid protection strategy (see Fig. 1 legend for details). To achieve random sequences in each coupling cycle, an equimolar mixture of all coded Fmoc-amino acids (cysteine was omitted to avoid disulphide bridges) was used for each cycle and coupling times were extended. Random peptides of increasing lengths were obtained by removing portions of the peptidylresin at the end of each synthesis cycle. Sequence analysis of the peptide mixture after 11 coupling cycles confirmed that the sequence was random. All amino acids (Trp was not quantitated) were present at each position, and the relative abundance of each amino acid was the same at all positions in the 11-mer (data not shown). Also, the amino-acid analysis of the water-soluble 7-mer peptide mixture showed that all amino acids were represented (Fig. 1 legend). Although the peptide does not contain equimolar amounts of each amino acid, we define this mixture to be random because all amino acids are found in all positions of the peptide. The chain-length dependence of the peptide-dependent ATPase activity of BiP was determined using random oligomers of different lengths. The initial rate of hydrolysis increases with increasing chain length from >4 to up to 7 residues (Fig. 1a), but longer peptides do not increase the rate of hydrolysis significantly. Peptide concentrations from 200 to 1,000 μ M gave the same result, with a maximal ATPase activity for seven residues.

Binding data correlate well with the ATPase experiments. BiP was incubated with ³H-labelled random peptides of different lengths, and peptides bound to BiP were isolated by rapid gel filtration. Again, peptide binding increased with chain length between 4 and 7 residues (Fig. 1b), but not with longer chains (up to 12). Peptides were tested at lower concentrations but gave similar results, ruling out the possibility that binding is saturated with the longer peptides. This binding was competitively inhibited by an excess of an unlabelled peptide KRQIYTDLEMNRLGK (in one-letter amino-acid code; termed A in our previous study¹²), which binds to BiP and stimu-

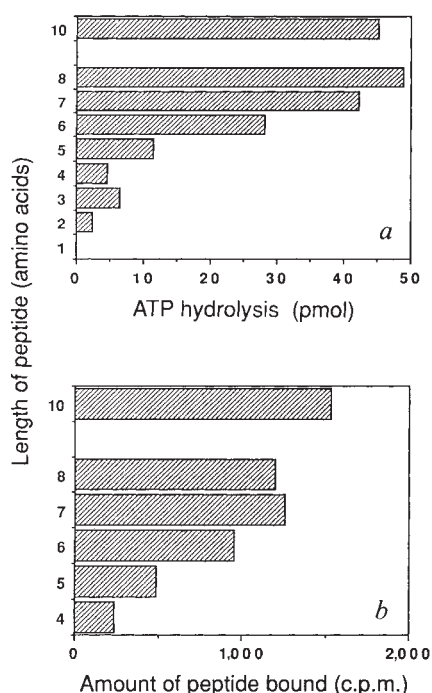


FIG. 1 Size requirement for peptide interaction with BiP: *a*, ATP hydrolysis; *b*, peptide binding.

METHODS. The random peptides were synthesized by solid-phase synthesis using the Fmoc-*t*-butyl amino-acid protection strategy³⁰. Synthesis was performed in an applied Biosystems model 430A peptide synthesizer on a *p*-hydroxymethylphenoxycetyl-polystyrene resin (0.2 mmol scale). Fmoc derivatives of Asp(O^tBu), Arg(Mtr), Glu(O^tBu), His(Trt), Lys(*t*Boc), Ser(*t*Bu), Tyr(*t*Bu) and Thr(*t*Bu) were used to incorporate Asp, Arg, Glu, His, Lys, Ser, Tyr and Thr, respectively. To achieve randomness in each position, an equimolar mixture of Fmoc derivatives of all coded amino acids was coupled in each synthesis cycle, with the exception of Fmoc-Arg(Mtr) and Fmoc-Ile, which were added in a 50% molar excess over the other derivatives (to compensate for a lower reactivity). To obtain the most random sequence, differences in reaction rates of the esters with the resin were offset using condensations (typically for 1 h at 25 °C) with equimolar amounts of resin and amino-acid derivatives. Coupling yields were ≥95%. At the end of each synthesis cycle, a portion of resin was removed, the peptides cleaved off the resin and deprotected in a mixture of trifluoroacetic acid/thioanisole/ethanedithiol (95:5:2) (*v/v/v*) for 3 h at 25 °C in a final volume of 1 ml. Peptides were extracted from the resin into neat trifluoroacetic acid (3 × 1 ml per 100 mg resin) and into dichloromethane (3 × 1 ml per 100 mg resin), the combined extracts concentrated to 0.5 ml (Savant SpeedVac concentrator) and the peptides precipitated with methyl-*t*-butyl ether (+4 °C for 12 h). Precipitates were washed with cold methyl-*t*-butyl ether, dissolved in water-acetonitrile (1:1, *v/v*) and freeze-dried. The 9-mer peptide mixture was unfortunately lost. Random peptides of fixed but defined lengths (*n*-mers) were solubilized by dissolving equimolar amounts in distilled water to make ~1 mM solutions. Insoluble peptides were removed by sedimentation at 15,000 *g* for 15 min. The remaining soluble peptides were decanted and stored. Concentrations were determined by the ninhydrin method³¹ after base hydrolysis. Repeating the centrifugation procedure did not remove measurable concentrations of peptide. The amino-acid analysis of the soluble 7-mer showed the composition to be (relative to Val) Ala 1.3, Arg 0.8, Asx 1.7, Glx 2.3, Gly 1.4, Ile 0.9, Leu 1.6, Lys 0.6, Met 0.3, Phe 0.7, Ser 1.2, Thr 0.9, Tyr 1.4, Val 1. The analysis was not corrected for losses of Ser, Tyr, Thr or Met. Peptide-dependent ATPase assays (*a*) were performed as before¹² using 350 μM and 700 μM concentrations of *n*-mer. Peptide binding (*b*) was assayed as described¹². Random peptides of defined lengths were labelled with tritium by reductive methylation and desalted over a G-10 Sephadex column (Pharmacia) to remove the free label^{12,32}. Labelled peptide mixtures were diluted with identical unlabelled peptides to achieve a specific activity of about 500 c.p.m. per pmol. The *n*-mer labelled peptides (4 and 8 nmol); were incubated with 2.5 μg BiP in 50 mM Tris-HCl, pH 7.5, 200 mM NaCl in a 50-μl volume for 15 min. The BiP-peptide complex was removed from the free peptide by spin-desalting through a 1-ml P-10 (Biorad) column at 900 r.p.m. in a GH-3.7 rotor in a Beckman GPC centrifuge for 2 min at 4 °C. Columns were prepacked under identical conditions. A factor was used to correct for small differences in specific activity.

lates hydrolysis of ATP. Also, excess unlabelled random peptides competitively inhibited the binding of ³H-labelled peptide A to BiP, with maximum inhibition at 7 residues and no further inhibition at increased chain lengths (data not shown). Thus, the random peptide- and the peptide A-binding sites on BiP are the same. Together these results argue for a binding site on BiP accommodating a peptide of up to seven residues in length.

Amino-acid preferences

Using the random-sequence heptameric peptide to fill the binding site of BiP with minimum degeneracy, we carried out a large-scale binding experiment (Fig. 2) with identical concentrations of peptide to those used in the binding experiment shown in Fig. 1 (80 μM), to determine the side chains preferred in binding. The BiP-peptide complex was removed from the total peptide mixture by rapid gel filtration, and was released upon denaturation of BiP. After removal from BiP, the peptide mixture was subjected to Edman degradation and the results compared with a control of with those obtained from the input soluble

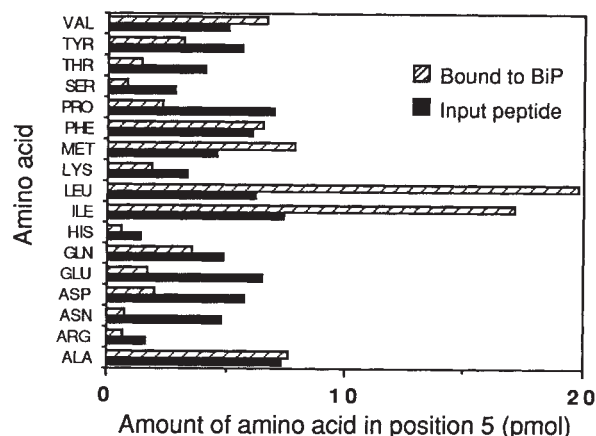


FIG. 2 Sequence analysis of the peptide ligand for position 5. This figure compares the bound peptide and the original random 7-mer passed through the reverse-phase resin, normalized for total amino-acid content at position 5.

METHODS. A large-scale binding reaction was set up with BiP and the 7-mer random peptide mixture: 57 μg of BiP was incubated with 50 nmol random 7-mer in 100 mM *N*-ethyl morpholine acetate, pH 7.5, in 600 μl for 30 min at 30 °C. Complexes were separated from unbound peptides by spin desalting over 1-ml P-10 columns (50 μl per column; see Fig. 1 legend). Eluates were pooled and the complexes denatured by adding trifluoroacetic acid to 0.1% final concentration. To separate bound peptides from BiP, the sample was made 20% in acetonitrile and passed through a C-18 reverse-phase Sep-pak cartridge (Waters). Under these conditions, over 95% of the random 7-mer flows through the cartridge while over 95% of BiP remains bound to the resin. As controls BiP was carried through all the reaction steps and peptide passed through the Sep-pak cartridge. The assay and peptide and protein controls were sequenced in an Applied Biosystems model 477A/120A protein sequencing system³³. The sample support disc was coated with 3 mg polybrene and prewashed using five FIL-1 cycles. The background contamination from solvents, support disc and sequencer chemicals was then determined (<1 pmol/PTH-amino acid) by running one modified³⁴ NORMAL-1 Edman degradation cycle before the samples were sequenced. Samples were applied to the disc in 50 μl each of water/acetonitrile/trifluoroacetic acid mixture (4:1:0.005 and 1:4:0.05, *v/v/v*, respectively). Gly and Trp were not quantitated as they co-eluted with contaminants, and Arg and Gln quantitation was subject to technical problems. Also, about 10% of Asn and 16% of Gln is converted during sequencing into Asp and Glu, respectively. A control with BiP alone passed through the spin-desalting and Sep-pak columns under identical conditions gave virtually no sequencing signal in the position corresponding to the peptide on the Sep-pak column. Moreover, the input peptide treated as described (without BiP) gave no signal in the position corresponding to BiP-peptide complex on the spin-desalting column, so there was no aggregation.

random heptameric peptide (see Fig. 2 legend). Equimolar amounts of each amino acid are not required as enrichment is expressed as a ratio of the bound to the input peptide mixture. Significant differences were observed (Fig. 2). The most striking difference was a consistent enrichment of the aliphatic amino acids at all positions in the bound peptide chains (Fig. 3a). Figure 2 presents the normalized data for position 5, and Fig. 3 (a–c) shows the relative abundance of each amino acid at each position in the Edman degradation. Values greater than unity indicate that an amino acid is selected to bind in that position, and less than one implies rejection. We encountered several technical problems related to the quantitation of amino acids. One artefact could arise from cumulative carry-over of phenylthiohydantoin (PTH)-amino acids into succeeding cycles of Edman degradation. This is not a major factor, however, because the recovery of amino acids from the peptide drops dramatically after the seventh (C-terminal) cycle of the 7-mer peptide (Fig. 3d). Another problem in interpreting the data could arise if the peptide were to shift to 'overhang' the binding pocket. But as shorter peptides bind with lower affinity, correctly positioned peptides will be enriched, making this unlikely.

The aromatic amino acids Phe and Tyr, although quite hydrophobic, are tolerated but not enriched (relative abundance ~ 1). Trp could not be determined because PTH-Trp coelutes with diphenylurea from the sequencer. Contamination of samples by free amino acids from buffers prevented us from obtaining data for position one, and for glycine. Polar uncharged amino-acid

side chains are excluded at all positions (Fig. 3b). Charged amino acids interact in a complex way with BiP (Fig. 3c): both negatively and positively charged amino acids are excluded from the binding site, with positive amino acids being more excluded, except for a slight enrichment at position 7. We do not know whether this enrichment represents a significant feature of the binding site or is merely a result of the favourable interaction of Lys or Arg with the carboxyl group at this terminal position—in a protein substrate this carboxyl would not be free. Proline is excluded from the binding site at all positions, indicating that BiP probably does not recognize proline-induced turns in secondary structure.

In summary, the binding site of BiP shows considerable specificity, with the degree of selection (as indicated for example by relative abundances) extending from as low as ~ 0.12 (position 3 for Ser) up to as high as ~ 3.5 (Leu at position 4), a dynamic range of about 30-fold. This factor corresponds to a binding energy of $\sim 2 \text{ kcal mol}^{-1}$ for each amino-acid side chain. Aliphatic side chains are preferred throughout the binding site, particularly in the core, but aromatic residues, although they are also hydrophobic, are excluded from much of the binding site for steric reasons.

We separated the random-sequence 7-mer into two fractions based on relative hydrophobicity and tested each for recognition by BiP. This separation gave 40% more Leu in the hydrophobic fraction (data not shown), which was markedly more efficient at stimulating the ATPase activity of BiP (Fig. 4). From

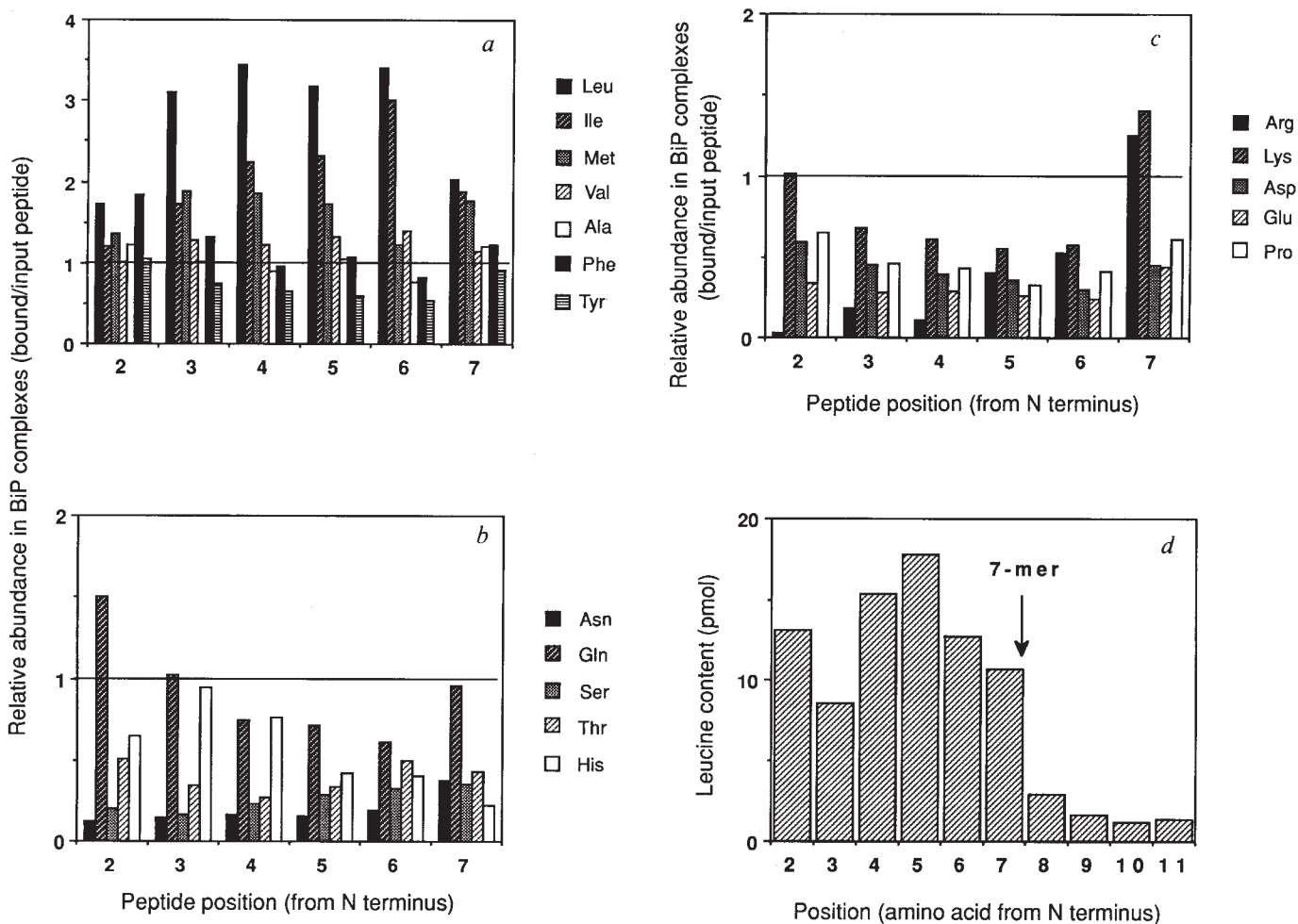


FIG. 3 Analysis of preferred BiP peptide substrate. The ratio of the bound to the input peptide is shown for all positions. A contaminant co-eluted with Gly during sequence analysis and was omitted. Buffer contaminants made analysis of position 1 impossible for all samples. A line is drawn at 1 on

the y axis to indicate the average value. Above the line denotes enrichment, below denotes exclusion for the indicated amino acid. a, Hydrophobic amino acids; b, hydrophilic amino acids; c, charged amino acids and proline; d, content of leucine as a function of position in the peptide.

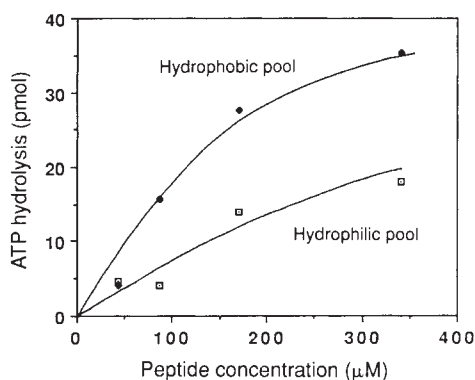


FIG. 4 Peptide-dependent ATPase assay using peptides differing in hydrophobic content. The random 7-mer was made 5% acetonitrile and 0.1% trifluoroacetic acid (v/v) and passed through a C-18 Sep-pak cartridge. Under these conditions, about 50% of the peptide passes through the cartridge (hydrophilic pool). The bound peptide pool was then eluted with 20% acetonitrile/0.1% trifluoroacetic acid (hydrophobic pool). These pools were dried in a Savant Speed-vac and made up to 0.5 mM. Peptide concentrations were determined by ninhydrin assay³¹ after base hydrolysis. Peptide-dependent ATPase was assayed as described¹².

extrapolation of the data in Fig. 4, we estimate that the K_m s of the two fractions differ by a factor of 5. We calculate that a random 7-mer would contain an average of 1.6 aliphatic residues in the hydrophobic fraction and ~ 1.1 aliphatic residues in the hydrophilic fraction.

Energy fingerprint

The relative abundance of an amino acid in the binding site reflects its distribution between free and bound states. The ratio of the relative abundances (A_1/A_2) of any two amino acids at the same position of the binding site is related to the difference ($\Delta\Delta G$) in their free energies of binding according to $\Delta\Delta G = \Delta G_2 - \Delta G_1 = -RT \ln(A_2/A_1) = -(RT \ln A_2 - RT \ln A_1)$, where R is the gas constant and T the temperature in degrees Kelvin. Figure 5a represents a composite of the data shown in Fig. 3 for positions 3–6, the 'core' of the peptide-binding site of BiP in the form of an energy 'fingerprint'. The y axis represents the relative abundance of each amino acid expressed on an energy scale, calculated as $\Delta G = -RT \ln A$, and the horizontal axis indicates its hydrophobicity¹³.

The energy fingerprint reveals a striking linear relationship between binding energy of the aliphatic residues and hydrophobicity. Moreover, amino-acid residues with unbranched side chains (Met) and side chains branched away from the peptide backbone (Leu) fall above the line; those with β -branched side chains fall below the line, showing that they bind more weakly. This means that flexibility of a side chain as well as its hydrophobicity are important in binding: the aromatics are well below the line, despite being the most hydrophobic, because they have the least flexible side chains.

The effect of substituting isoleucine at position 3 of T4 lysozyme on the thermostability of this enzyme has been investigated¹⁴; this position corresponds to the hydrophobic core of the N-terminal region. The differences in the free energy of binding of the substituted amino acids to this pocket, which normally houses an aliphatic residue in the interior of the protein, are replotted in our Fig. 5b (their Fig. 2a). This energy fingerprint of a pocket that binds an aliphatic residue to the interior of a folded protein is qualitatively indistinguishable from the fingerprint of the core of the peptide-binding site of BiP (Fig. 5a). The slope of the line indicates the fraction of hydrophobic energy potentially available in binding (probably related to solvent inaccessibility after binding) and is greater for T4 lysozyme (0.87) than for BiP (0.51).

Discussion

We have found that the peptide-binding site of BiP is filled by a stretch of only 7 residues, verifying that hsp70 members bind single-strand polypeptides¹². This property enables hsp70s to coat newly translated or translocated polypeptide chains and prevents them from folding prematurely. The turnover time of the BiP peptide-dependent ATPase (5 min) can keep pace with the speed of translation and translocation^{5,12}. Although the peptide-binding site prefers to bind aliphatic side chains, as typically found inside folded proteins^{15,16}, it will tolerate binding of polar and charged residues. BiP can therefore accommodate varied segments of linear chains destined to assume different folded forms. For example, the arrangement of hydrophobic residues along a chain is different for an α -helix and a β -sheet, which have globular protein periodicities of 3 or 4 residues and 2 residues respectively¹⁷. A span of 7 amino acids containing an average of only 1.6 aliphatic residues (the hydrophobic fraction of Fig. 4) binds productively to BiP. In a typical globular protein with 20% aliphatic content¹⁸, we calculate that a 7-amino-acid stretch will occur every 16 residues along the chain on average. Thus, potential binding sites for hsp70s are very common. Such stretches will be sterically accessible to hsp70s when the chain is in an extended form, as during translation and translocation, allowing binding of polypeptides that would otherwise exist only transiently in extended states.

Hsp70s have many chaperoning functions^{5,19,20}, all of which require that they recognize incompletely folded and unfolded chains but ignore native conformations. Our results give some indication of how this recognition is achieved. The virtually indistinguishable energy fingerprints of the peptide-binding site of an hsp70 and of an aliphatic residue inside a folded protein indicate that hsp70 can substitute for the hydrophobic interior of a folded globular protein. This explains how hsp70s might recognize unfolded structures, and how molecular chaperones collectively define a native structure for the cell. Any 'folded' protein that still binds hsp70s will probably be nonfunctional, so correct cellular proteins must remain folded in the presence of hsp70s with their hsp70 binding sites sequestered to the interior.

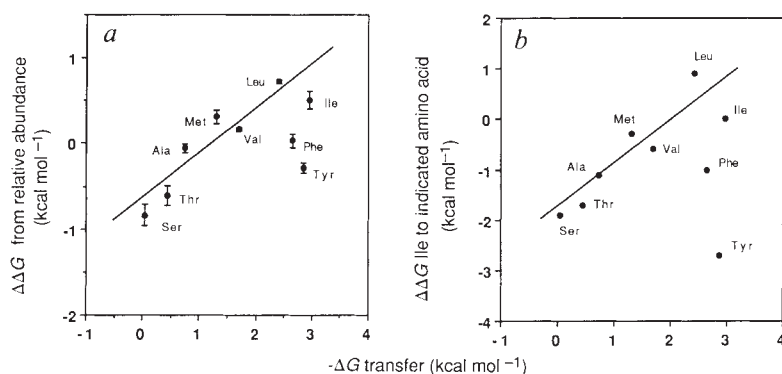


FIG. 5 Enrichment of amino acids versus hydrophobicity. *a*, Log of the enrichment of the hydrophobic amino acids at positions 3 to 6 is a measure of the free energy change on binding ($-\Delta G = 1.4 \log K_{eq}$). The hydrophobicity of the residue is plotted as the free energy change on transfer from ethanol to water^{13,35}. Vertical bars indicate the standard deviation above and below the mean value. *b*, Reproduction of Fig. 2a of ref. 14 showing the free energy of stabilization of T4 lysozyme mutated in position 3 plotted against hydrophobicity of individual amino acids.

An N-terminal 44K fragment of cytoplasmic hsp70 produced by protease cleavage contains the ATP-binding domain but cannot bind the polypeptide clathrin²¹, suggesting that the peptide-binding region of hsp70 may lie in the cleaved C-terminal region. A loose homology between this C-terminal region of hsp70 and the probable peptide-binding site on the *Xenopus* major histocompatibility complex class I (MHC class I) molecule has been discovered²². A structure of the peptide-binding pocket of hsp70 was proposed by analogy with the structure of human MHC class I (ref. 23). The minimum peptide chain length capable of binding to

class I seems to be 8 or 9 amino acid residues²⁴⁻²⁶. But the floor of the proposed peptide-binding groove of hsp70 contains one less β -sheet, so if an extended conformation of the bound peptide is assumed for both the MHC molecule and hsp70, this result is consistent with our conclusion that the binding site of hsp70 can accommodate 7 residues of a protein.

Our stochastic approach should prove useful in analysing other cases of broad specificity for which existing methods²⁷⁻²⁹ have shortcomings, such as the peptide binding of MHC molecules. □

Received 29 July; accepted 13 September 1991.

- Lippincott-Schwartz, J., Bonifacino, J. S., Yuan, L. C. & Klausner, R. D. *Cell* **54**, 209-220 (1988).
- Goldberg, A. L. & St John, A. C. *Rev. Biochem.* **45**, 747-803 (1976).
- Ostermann, J., Horwich, A. L., Neupert, W. & Hartl, F. U. *Nature* **341**, 125-30 (1989).
- Bole, D. B., Hendershot, L. M. & Kearney, J. F. *J. Cell Biol.* **102**, 1558-1566 (1986).
- Rothman, J. E. *Cell* **59**, 591-601 (1989).
- Gething, M.-J., McCammon, K. & Sambrook, J. *Cell* **46**, 939-950 (1986).
- Chirico, W. J., Waters, M. G. & Blobel, G. *Nature* **332**, 800-805 (1988).
- Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. *Nature* **332**, 800-804 (1988).
- Kang, P.-J. *et al. Nature* **348**, 137-143 (1990).
- Beckmann, R. P., Mizzen, L. A. & Welch, W. J. *Science* **248**, 850-853 (1990).
- Palleros, D. R., Welch, W. J. & Fink, A. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5719-5723 (1991).
- Flynn, G. C., Chappell, T. G. & Rothman, J. E. *Science* **245**, 385-390 (1989).
- Nozaki, Y. & Tanford, C. *J. Biol. Chem.* **246**, 2211-2217 (1971).
- Matsumura, M., Becktel, W. J. & Matthews, B. J. *Nature* **334**, 406-410 (1988).
- Miller, S., Janin, J., Lesk, A. M. & Chothia, C. *J. molec. Biol.* **196**, 641-656 (1987).
- Lim, W. A. & Sauer, R. T. *Nature* **339**, 31-36 (1989).
- Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 140-144 (1984).
- Kapper, M. H. *Biochem. biophys. Res. Commun.* **78**, 1018-1024 (1977).

- Lewis, M. J. & Pelham, H. R. B. *EMBO J.* **4**, 3137-3143 (1985).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Chappell, T. G., Konforti, B. B., Schmid, S. L. & Rothman, J. E. *J. Biol. Chem.* **262**, 746-751 (1987).
- Flajnik, M. F., Canel, C. C., Kramer, J. & Kasahara, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 537-541 (1991).
- Flajnik, M. F., Canel, C. C., Kramer, J. & Kasahara, M. *Immunogenetics* (in the press).
- Rotzschke, O. *et al. Nature* **348**, 252-254 (1990).
- Schumacher, T. N. M. *et al. Nature* **350**, 703-706 (1991).
- Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. *Nature* **351**, 290-296 (1991).
- Devlin, J. J., Panganiban, L. C. & Devlin, P. E. *Science* **249**, 404-406 (1990).
- Scott, J. K. & Smith, G. P. *Science* **249**, 386-390 (1990).
- Fodor, S. P. A. *et al. Science* **251**, 767-773 (1991).
- Chang, C. D. & Meienhofer, J. *Int. J. Pept. Protein Res.* **11**, 246-249 (1975).
- Cresfield, A. M., Stein, W. H. & Moore, S. *J. Biol. Chem.* **238**, 618-621 (1963).
- Tack, B. F., Dean, J., Eilat, D., Lorenz, P. & Schechter, A. N. *J. Biol. Chem.* **255**, 8842-8847 (1980).
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. Biol. Chem.* **256**, 7990-7997 (1981).
- Pohl, J., Pereira, H. A., Martin, N. M. & Spitznagel, J. K. *FEBS Lett.* **272**, 200-204 (1990).
- Tanford, C. *J. Am. Chem. Soc.* **84**, 4240-4247 (1962).

ACKNOWLEDGEMENTS. We thank M. Hecht for discussion, and D. Wilson, R. Baldwin, B. Matthews, M. Wiedmann and M. G. Waters for comments on the manuscript. Supported in part by the NIH (J.E.R.) and the Damon Runyon-Walter Winchell Cancer Fund (G.C.F.).

LETTERS TO NATURE

High-energy background radiation from quiescent γ -ray burst sources

Jan van Paradijs

Astronomical Institute 'Anton Pannekoek', University of Amsterdam and Center for High Energy Astrophysics, NIKHEF-H, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA

COSMIC γ -ray bursts (GRBs) were discovered¹ 20 years ago, but remain puzzling. The detection²⁻⁴ of cyclotron harmonics in the spectra of some GRBs suggests an origin in magnetized neutron stars, but there is no consensus on the energy source or the mechanism that generates the γ -rays. Statistical analysis⁵ of the spatial distribution of GRBs indicates an origin in our Galaxy, and it has been proposed that they are caused by thermonuclear flashes in accreted material on the surface of strongly magnetized neutron stars⁶⁻⁹. I point out here that, in this model, accretion during the quiescent intervals between bursts will generate a diffuse background emission from what may be a very large number of objects, too faint to be individually detected. Because the quiescent emission is related directly to the observed intensity and frequency of GRBs throughout the Galaxy, measurement of the diffuse background in the appropriate energy band may provide a test of the thermonuclear-flash model which is independent of the distance scale to GRB sources, the possible emission anisotropy of both burst and quiescent radiation, and the accretion rates and recurrence times of individual sources.

The thermonuclear-flash model has been successful in describing the properties of X-ray bursts emitted by the weakly magnetized neutron stars in low-mass X-ray binaries¹⁰, and the large differences between these X-ray bursts and GRBs are supposed to be related to large differences in both the strength of the

magnetic field of the neutron star, and the accretion rate onto it¹¹. If GRBs are the result of thermonuclear flashes on the surface of a neutron star, this neutron star must be a source of 'quiescent' emission caused by the accretion. The average accretion luminosity is larger than the (long-term) average luminosity in GRB by a factor that is equal to the ratio, α , of the gravitational binding energy (at the neutron star surface) to the available nuclear energy, both taken per unit mass. The latter depends on the type of nuclear fuel (for example hydrogen or helium), and on the fraction of the fuel that has been steadily burnt during the 'quiescent' time intervals between GRB. For a 'canonical' neutron star with a mass of $1.4 M_{\odot}$ and a radius of 10 km, $\alpha > 20$ for hydrogen flashes, and $\alpha > 10^2$ for helium flashes; the only detailed calculation⁶ of thermonuclear flashes on magnetic neutron stars of which I am aware leads to $\alpha \approx 120-140$.

For a population of GRB sources, one therefore expects to observe an average intensity (see ref. 12) due to accretion which is α times the average intensity observed in the GRB. This provides a method that may in principle allow us to test the thermonuclear-flash model by comparing the expected accretion intensity (and its angular distribution on the sky) with an observed diffuse background. Because the specific intensity does not change along a line of sight¹² (unless absorption occurs), this method is independent of the distance scale to GRB sources. Furthermore, it is independent of the possible anisotropy of the burst emission and of the quiescent accretion luminosity, as these are not expected to be correlated with the source location. Finally, as the argument is based on an energy ratio for a population of accreting sources, the method is also independent of the accretion rate of an individual source and (coupled to this) the recurrence time of GRB. Uncertainties in all these quantities strongly affect a similar method applied to individual GRB sources^{13,14}. (Note that neutron stars in general, regardless of whether they are GRB sources or not, are expected to accrete from the interstellar medium; therefore, the background radiation due to this accretion may have a broader application than GRB sources only).

The observed sky distribution of the GRBs is highly isotropic (see for example refs 5, 15). As discussed extensively^{16,17}, if the GRB sources are a galactic population of neutron stars, this isotropy is hard to reconcile with the observed cumulative size-frequency distribution (the 'log $N(>S)$ -log S curve', where S is the fluence of the GRB and N the frequency). Below $S \approx 10^{-5}$ erg cm⁻², this curve shows a flattening with respect to the power law [$N(>S) \propto S^{-3/2}$] expected for a homogeneous spherical source distribution. It has been suggested¹⁸⁻²¹ that this flattening may be caused by incomplete sampling of low-fluence bursts (particularly those which last relatively long) because of instrumental bias. Preliminary results obtained²² with the BATSE experiment on board the Gamma Ray Observatory indicate, however, that the flattening is probably real (see also refs 23, 24).

Roughly one burst per day was detected²² by BATSE (whose detections are expected to be complete down to a fluence $S \approx 3 \times 10^{-7}$ erg cm⁻²; see ref. 25) during the first three months of its operation. From this and from previous results²⁶ for bright GRBs ($S > 10^{-5}$ erg cm⁻²), we will approximate the cumulative fluence distribution as follows: for the bright bursts ($S > 10^{-5}$ erg cm⁻²), $N(>S) = 2.5 (S/S_0)^{-3/2}$ yr⁻¹, and for weak bursts ($S < 10^{-5}$ erg cm⁻²), $N(>S) = 20 (S/S_0)^{-1/2}$ yr⁻¹; here $S_0 = 10^{-4}$ erg cm⁻².

With this cumulative fluence distribution we find for the average specific intensity I_{GRB} (in GRBs with fluences down to $S = 10^{-8}$ erg cm⁻²) a value $I_{\text{GRB}} = 8 \times 10^{-12}$ erg cm⁻² s⁻¹. As some residual instrumental bias may be present in the preliminary BATSE results, this number is to be considered a lower limit. Corresponding to this one expects an average background intensity I_{acc} (possibly due to a very large number of accreting quiescent GRB sources, each of which is too faint to be detectable) in excess of $\sim 2 \times 10^{-10}$ or $\sim 8 \times 10^{-10}$ erg cm⁻² s⁻¹ sr⁻¹ (for hydrogen flashes and helium flashes, respectively).

It is unclear, *a priori*, in which energy band the bulk of the accretion energy will be emitted. If we may take the results obtained for known accreting strongly magnetized neutron stars (binary X-ray pulsars; see ref. 27 for a recent review) as a guideline, we would expect the bulk of the accretion power to emerge in the range between ~ 1 and ~ 100 keV. The argument presented here is, however, independent of the energy band in which the accretion power is radiated, as long as this power is not absorbed by the interstellar medium (as would be the case if it were emitted in the soft X-ray or EUV bands).

Using the results on the diffuse X-ray background collected by Cowsik and Kobetich²⁸, we find that its specific intensity I_{XRB} is $\sim 7 \times 10^{-8}$, $\sim 1.1 \times 10^{-7}$, and $\sim 6 \times 10^{-8}$ erg cm⁻² s⁻¹ sr⁻¹, in the energy bands 1-10 keV, 10-100 keV and 100-1,000 keV, respectively. It thus seems that, under the above assumptions, the accretion required to fuel the thermonuclear flashes (which in our hypothesis cause the GRB's) would provide only a small contribution to the X-ray background.

For a galactic disk-like distribution of GRB sources, one expects that below some fluence level the sky distribution of GRB's becomes anisotropic; anisotropy is therefore also expected for the background radiation caused by the accreting quiescent GRB sources. Evidence for anisotropy in the diffuse X-ray background, in the form of a galactic component whose specific intensity decreases with galactic latitude, has been found²⁹⁻³¹ from scanning data obtained with Ariel V, HEAO-1 and Uhuru. Taking the Ariel V results²⁹, which are in reasonable quantitative agreement with the other results mentioned, we find that in the energy range ~ 2 -10 keV, this diffuse galactic component contributes $\sim 4\%$ to the diffuse X-ray background, corresponding to an average specific intensity of $\sim 3 \times 10^{-9}$ erg cm⁻² s⁻¹ sr⁻¹. This is close to the limits derived above on I_{acc} due to accreting quiescent GRB sources.

In view of the obvious quantitative uncertainties in the above analysis, these comparisons of expected and observed background signals are not yet very meaningful. They serve, however,

to show that in the future a similar, more detailed analysis of the emission expected from the accreting quiescent GRB sources may in principle provide a test of the thermonuclear-flash model of GRB.

Prerequisites for such an analysis are a knowledge of the photon spectrum emitted by an accreting strongly magnetized star, at accretion rates many order of magnitudes below those encountered in binary X-ray pulsars, and unbiased fluence measurements and angular distributions for GRBs down to a fluence level of $\sim 10^{-7}$ erg cm⁻². Observations with the BATSE experiment²³ on board the Gamma Ray Observatory are expected to provide these distributions in the near future. $\square$

Received 23 May; accepted 9 September 1991.

- Klebesadel, R. W., Strong, I. B. & Olson, R. A. *Astrophys. J.* **182**, L85-88 (1973).
- Mazets, E. P. et al. *Nature* **290**, 378-382 (1981).
- Murakami, T. et al. *Nature* **335**, 234-235 (1988).
- Fenimore, E. E. et al. *Astrophys. J.* **335**, L71-74 (1988).
- Atteia, J. L. et al. *Nature* **351**, 296-298 (1991).
- Woosley, S. E., Wallace, R. K. *Astrophys. J.* **258**, 716-732 (1982).
- Woosley, S. E. in *High Energy Transients in Astrophysics*, AIP Conf. Proc. **115** (ed. Woosley, S. E.) 485-511 (American Institute of Physics, 1984).
- Hameury, J. M., Bonazzola, S., Heyvaerts, J. & Ventura, J. *Astr. Astrophys.* **111**, 242-251 (1982).
- Hameury, J. M., Bonazzola, S. & Heyvaerts, J. *Astr. Astrophys.* **121**, 259-264 (1983).
- Lewin, W. H. G. & Joss, P. C. in: *Accretion-driven Stellar X-ray Sources* (eds Lewin, W. H. G. & van den Heuvel, E. P. J.) 41-115 (Cambridge University Press, 1983).
- Hameury, J. M. & Lasota, J. P. in *Gamma Ray Bursts*, AIP Conf. Proc. **141**, (eds Liang, E. P. & Petrosian, V.) 177-188 (American Institute of Physics, 1986).
- Mihalas, D. *Stellar Atmosphere*, 2nd Ed. Ch. 1 (Freeman, San Francisco, 1978).
- Pizzichini, G. et al. *Astrophys. J.* **301**, 641-649 (1986).
- Boer, M. et al. *Astr. Astrophys.* **202**, 117-123 (1988).
- Hartmann, D. & Epstein, R. I. *Astrophys. J.* **346**, 960-966 (1989).
- Higdon, J., Meegan, C. & Cline, T. in *Gamma Ray Bursts*, AIP Conf. Proc. **141**, (eds Liang, E. P. & Petrosian, V.) 23-32 (American Institute of Physics, 1986).
- Hurley, K., Cline, T. & Epstein, R. I. in *Gamma Ray Bursts*, AIP Conf. Proc. **141**, (eds Liang, E. P. & Petrosian, V.) 33-38 (American Institute of Physics, 1986).
- Higdon, J. C. & Lingenfelter, R. E. in *High Energy Transients in Astrophysics*, AIP Conf. Proc. **115**, (ed. Woosley, S. E.) 568-577 (American Institute of Physics, 1984).
- Mazets, E. P. *Proc. 19th Int. Cosmic Ray Conf.* (eds Jonas, C. F. et al.) 424-431 (NASA Washington DC 1986).
- Paczynski, B. & Long, K. *Astrophys. J.* **333**, 694-699 (1988).
- Schmidt, M., Higdon, J. C. & Hueter, G. *Astrophys. J.* **329**, L85-87 (1988).
- Chipman, E. *Gamma Ray Obs. Science Rep.* No. 23 (25 July 1991).
- Meegan, C. A., Fishman, G. J. & Wilson, R. B. *Astrophys. J.* **291**, 479-485 (1985).
- Ambruster, C. W. & Wood, K. S. *Astrophys. J.* **311**, 258-274 (1986).
- Fishman, G. J. et al. in *Proc. Gamma Ray Observatory Science Workshop* (ed. Johnson, W. N. (Naval Research Laboratory, Washington, DC 3.47-62 (1989).
- Jennings, M. *Astrophys. J.* **333**, 700-718 (1988).
- Nagase, F. *Publ. astr. Soc. Japan* **41**, 1-75 (1990).
- Cowsik, R. & Kobetich, E. J. *Astrophys. J.* **177**, 585-592 (1972).
- Warwick, R. S., Pye, J. P., Fabian, A. C. *Mon. Not. R. Astr. Soc.* **190**, 243-260 (1982).
- Iwan, D. et al. *Astrophys. J.* **260**, 111-123 (1982).
- Protheroe, R. J., Wolfendale, A. W. & Wdowczyk, J. *Mon. Not. R. Astr. Soc.* **192**, 445-454 (1982).

ACKNOWLEDGEMENTS. I thank B. Paczynski for his comments on this paper. This work was supported by NATO.

Eclipses of the ablating binary pulsar PSR1744-24A

S. E. Thorsett* & D. J. Nice

Joseph Henry Laboratories and Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

PSR1744-24A, a binary pulsar in the globular cluster Terzan 5, is eclipsed for up to half of each 109-minute orbit^{1,2}. This is a much longer duration than its 0.09-solar-mass companion star could cause if it were confined within its Roche lobe, and has led to the suggestion that PSR1744-24A is ablating material from its companion, producing an extensive stellar wind that eclipses the radio signal. We have measured the pulsar light curve at a frequency of 1.67 GHz, and find that the eclipse is not total: maximum attenuation in some cases is only about 70 per cent. The pulsar signal is substantially delayed during the eclipse. We model these observations by means of free-free absorption and

* Present address: Radio Astronomy, 105-24, California Institute of Technology, Pasadena, California 91125, USA.

dispersion in an ionized wind, processes which also explain the frequency dependence of the eclipse duration³. A simplified model predicts a wind density at the Roche surface of $\sim 6 \times 10^7$ electrons cm^{-3} and an electron temperature in the wind of $(3.6-15) \times 10^3$ K, although some theoretical difficulties remain. Such a wind is unlikely to ablate the companion star entirely in less than a Hubble time, so that in this one case at least, ablation may not ultimately form an isolated millisecond pulsar.

In 1987, Ruderman *et al.*^{4,5} suggested a 'bootstrap' mechanism for rapid evaporation of light companions in low-mass X-ray binaries. The neutron star is spun up during a self-sustaining episode of mass transfer, which is quenched suddenly as the secondary sinks below its Roche lobe, leaving behind a radio pulsar with a very light companion. The continued evaporation and eventual elimination of this companion by high-energy pulsar radiation was proposed as an attractive means of producing isolated millisecond pulsars, such as PSR1937+21.

Two such examples of binary pulsars that are evaporating their companions are now known. Both PSR1957+20 and PSR1744-24A are in close eclipsing binary systems in which the eclipses last too long to be caused by material confined within the companion's Roche lobe. The implication in each case is that the companion is losing mass, presumably ablated from its surface by energetic pulsar radiation. Because the future of the binary system depends directly on the evaporation rate, the nature of this stellar wind has been the focus of much attention³⁻¹⁰.

We observed PSR1744-24A at several epochs between March 1990 and April 1991 using the Green Bank 43-m telescope and the Very Large Array (VLA). Here we report on eclipse observations made at both sites near 1.67 GHz. Green Bank observations at other frequencies, as well as long-term timing results, will be discussed elsewhere.

At Green Bank, a digital fast-Fourier-transform spectrometer was used to divide a 40-MHz bandwidth into 512 channels of 78 kHz each; at the VLA, an analog filter bank provided fourteen 2-MHz channels. At each site, the received signal was averaged coherently at the topocentric pulsar period; integration times varied from 30 to 300 s. The channels were aligned to compensate for dispersive delays, and the orthogonal polarizations added, to produce a single average pulse profile for each integration. Standard techniques were used to measure the topocentric pulse arrival times, reduce these to effective arrival times at the Solar System barycentre and fit the resulting data to a binary pulsar timing model.

TABLE 1 Attenuation of signal during eclipse

Date (MJD)*	Optical depth	Date (MJD)*	Optical depth
48190	1.61 (12)	48363	1.33 (5)
48225	1.78 (14)	48364	1.28 (8)
48268	>2.03	48369	1.19 (9)

* Modified Julian date.

Our initial observations of PSR1744-24A in March and April 1990 found that the pulsed signal disappeared for 30-50% of the orbit, with longer eclipses at lower observing frequencies². The signal was seen to fade slowly in or out at eclipse edges, and there were some indications that the pulses were occasionally delayed. We were surprised to find that on several later occasions the pulsar signal, although strongly attenuated, was not completely eclipsed. Figure 1a shows the light curve measured at the VLA from data taken in April 1991, the epoch showing the least attenuation of signal during eclipse. At the point of inferior conjunction, the signal strength is reduced by a factor of ~ 3.3 , corresponding to an optical depth $\tau \approx 1.2$. The list of optical depths at various epochs in Table 1 demonstrates that the amount of attenuation is highly variable. (The optical depths were calculated by comparing signals between orbital phases 0.15 and 0.35 to generate an on-eclipse strength, and 0.50 and 1.00 for an off-eclipse strength.) This variability may explain why the signal was not detected within eclipse during earlier observations; the initial Green Bank observations were also hampered, however, by significantly lower sensitivity than is available at the VLA. The variability in attenuation and eclipse length and the occasional disappearances of pulsed signals for entire orbits^{1,2} strongly suggest an unstable companion wind.

In addition to the signal attenuation, the pulses are substantially delayed during eclipse, by an average of ~ 500 μs at 1.67 GHz during inferior conjunction (Fig. 1b). Sporadic delays have previously been reported for PSR1744-24A at eclipse edges and after anomalous signal disappearances^{1,2}, but our observations show a delay curve throughout the eclipsing region.

These new observations of signal delays and attenuation provide challenging tests for the mechanisms that have been proposed to explain the radio eclipses of PSR1744-24A and PSR1957+20⁶⁻⁸, especially when combined with the strong dependence of eclipse length on observing frequency² and the

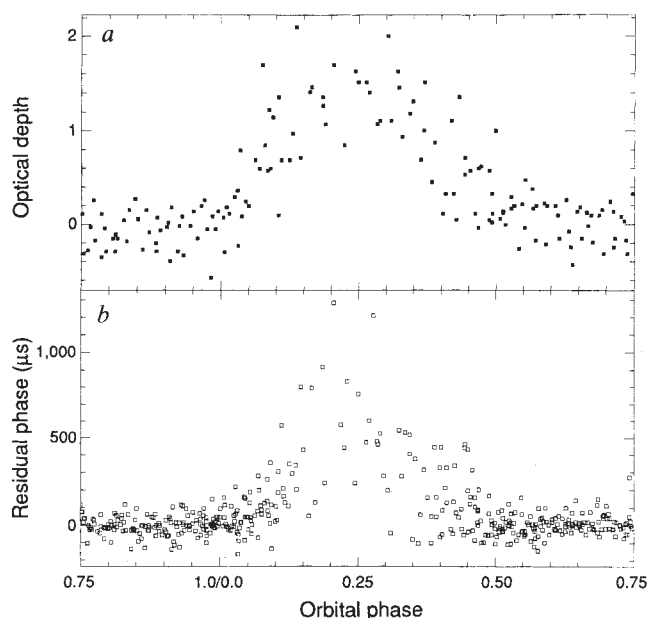


FIG. 1 a, The light curve of PSR1744-24A. Inferior conjunction occurs at orbital phase 0.25. The optical depth τ is calculated relative to the mean off-eclipse flux. Only VLA data from MJD 48363 are plotted. b, Residual pulse arrival times after subtraction of the pulsar orbital model. Plot includes all data from Green Bank and the VLA.

extreme symmetry of the eclipses around inferior conjunction. A full discussion of the various models must be left to a later paper. In summary, however, we believe that the most compelling arguments favour models based on free-free absorption in an ionized medium, despite some serious theoretical difficulties with the production of, and conditions in, the absorbing plasma.

The frequency dependence of the eclipse length first lent credence to free-free absorption mechanisms^{3,9,10}. The observed time delays strengthen these models, as radio signals are delayed by propagation through an ionized medium. Using the maximum delay of 500 μ s at our observing frequency of 1.67 GHz gives an electron column density of $\sim 1.0 \times 10^{18} \text{ cm}^{-2}$ at the closest approach of the line of sight to the pulsar-companion axis. A rough estimate of the electron density comes from dividing this figure by the pulsar-companion separation, $0.85 R_{\odot}$, giving $n_e \approx 2 \times 10^7 \text{ cm}^{-3}$. If we further assume a uniform electron distribution, the plasma temperature required to produce the observed optical depth in a thermal free-free absorption model¹¹ is $\sim 7,000 \text{ K}$.

A more refined (although still simplified) model was suggested by Rasio *et al.*⁹, in which the wind is spherically symmetric about the companion and the electron density drops as $n_e(r) \propto r^{-2}$. Although these are clearly idealizations, the eclipse symmetry does provide some support for these assumptions. (A model incorporating emission of the wind from only the half of the companion nearest the pulsar gives nearly the same results.)

The model incorporates the inclination of the orbit to the line of sight, which is unknown, except that it must be large enough for the line of sight not to intersect the companion, yet small enough to allow the eclipse modulation to be seen. We assume an impact parameter at inferior conjunction of twice the Roche lobe radius, $2r_R \approx 0.3 R_{\odot}$, noting that our conclusions will be rather insensitive to this assumption.

In this model, the free electron density is again calculated from the observed delays during eclipse: $n_e(r) \approx 6 \times 10^7 \text{ cm}^{-3} (r/r_R)^{-2}$. In the free-free absorption model, the electron temperature can be calculated in two extreme cases⁹: an isothermal wind, $T(r) = T_R$, and an adiabatically cooled wind, $T(r) = T_R(r/r_R)^{-4/3}$. From the observed optical depth at inferior conjunction, the required electron temperature is $T \approx 3,600 \text{ K}$ for the isothermal case and $T(r) \approx 15,000(r/r_R)^{-4/3} \text{ K}$ for the adiabatic case. The similarity of these results to our previous rough calculation suggests that similar general results can be expected from less idealized assumptions.

The low temperatures predicted are the primary difficulty with free-free models. For thermal ablation to occur, the companion's surface must be extremely hot. For a $0.1 M_{\odot}$, $0.1 R_{\odot}$ companion, the 'escape temperature' for hydrogen is roughly $6 \times 10^6 \text{ K}$, and it is far from clear how to cool the material rapidly to the observed temperatures. A complete free-free model must therefore include either an efficient cooling mechanism, or a nonthermal source for the companion wind, such as Roche lobe overflow.

Assuming this objection can be overcome, we can use the observed electron density to estimate the rate of mass loss from the companion, and hence how long it will last. If the wind has velocity v and ionization fraction ξ (which is frozen in after ablation, as the wind is far from ionization-recombination equilibrium¹⁰), then the evaporation timescale is

$$\left(\frac{m}{\dot{m}}\right) = \left(\frac{m\xi}{4\pi r_R^2 \rho_R v}\right) \approx 700\xi \left(\frac{v}{v_{\text{orb}}}\right)^{-1} \text{ Gyr}$$

where $v_{\text{orb}} = 6 \times 10^7 \text{ cm s}^{-1}$ is the companion's orbital velocity. For ionization fractions greater than a few per cent and any reasonable values of v , this time is greater than the Hubble time, indicating that ablation alone may be insufficient to leave PSR1744-24A as an isolated millisecond pulsar. $\square$

Received 20 May; accepted 9 September 1991.

1. Lyne, A. G. *et al. Nature* **347**, 650-652 (1990).
2. Nice, D. J., Thorsett, S. E., Taylor, J. H. & Fruchter, A. S. *Astrophys. J.* **361**, L61-L63 (1990).
3. Rasio, F. A., Shapiro, S. L. & Teukolsky, S. A. *Astr. Astrophys.* **241**, L25-L28, (1991).
4. Ruderman, M., Shaham, J. & Tavani, M. *Astrophys. J.* **336**, 507-518 (1989).
5. Ruderman, M., Shaham, J., Tavani, M. & Eichler, D. *Astrophys. J.* **343**, 292-312 (1989).
6. Fruchter, A. S. *et al. Astrophys. J.* **351**, 642-650 (1990).
7. Michel, F. C. *Nature* **337**, 236-238 (1989).
8. Eichler, D. *Astrophys. J.* **370**, L27-L30 (1991).
9. Rasio, F. A., Shapiro, S. L. & Teukolsky, S. A. *Astrophys. J.* **342**, 934-939 (1989).
10. Wasserman, I. & Cordes, J. M. *Astrophys. J.* **333**, L91-L94 (1988).
11. Rybicki, G. B. & Lightman, A. P. *Radiative Processes in Astrophysics* (Wiley, New York, 1979).

ACKNOWLEDGEMENTS. We thank M. McKinnon for observational assistance at the VLA and J. Taylor for helpful comments on the manuscript. The NRAO is operated by Associated Universities, Inc., under contract from the NSF. This research was supported by the NSF and the U.S. Department of Education.

Observation of quantum supershells in clusters of sodium atoms

J. Pedersen*, S. Bjørnholm*, J. Borggreen*,
K. Hansen*, T. P. Martin† & H. D. Rasmussen*

* The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

† Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, D-7000 Stuttgart 80, Germany

ATOMIC clusters of sodium and other simple metals are known to exhibit a shell structure, giving rise to enhanced stability at certain 'magic numbers' of constituent atoms¹⁻³. Balian and Bloch have shown⁴ that such shells are the likely result of particularly stable electronic structures: electrons in a spherical cavity (approximating the potential in the clusters) follow semiclassical triangular or square orbits, leading to a shell structure similar to that in atoms⁵, and stable configurations occur at magic numbers proportional to the cube root of the number of electrons. Balian and Bloch⁴ also predicted that the existence of both triangular and square orbits, with slightly different periodicities of their magic numbers, should lead to a 'quantum beating' effect that imposes a low-frequency envelope on the periodic variation in cluster stability with increasing size, in effect creating an additional 'supershell' structure. Here we report the observation of this supershell effect in sodium clusters with up to 3,000 constituent atoms.

The semiclassical theory of Balian and Bloch⁴ was extended and applied to idealized spherical sodium droplets with somewhat diffuse potential walls⁶. A quantum beat was found in the shell structure, with a pronounced minimum being predicted around size $N = 1,000$. The semiclassical results were compared to and found to agree fairly closely with the eigenvalue spectrum obtained by solving the stationary Schrödinger equation for independent electrons in the same potential. In hydrogen there is a one-to-one relation between a classical orbit and the corresponding quantum eigenstate. Both pictures predict the well-known shell structure in hydrogen. With potentials like those of a spherical metal cluster, the one-to-one correspondence between individual classical and quantum 'orbits' is lost. Nevertheless, the bunching of quantum levels into shells can still be understood^{4,6} in terms of the system of closed orbits (polygons with more or less rounded corners) available to a classical particle moving in the same potential. From this description it becomes possible to see that the quantum beat mode results from an interference between classical triangular and square orbits, because these emerge as the dominant and equally important contributors to the quantum shell structure^{4,6}.

Experiments have shown that the spherical mean field assumption used in the shell structure calculations is valid for sodium clusters extending in size beyond 600 (ref. 7), possibly even to $N \leq 1,400$ (ref. 8). A recent laser warming experiment⁹

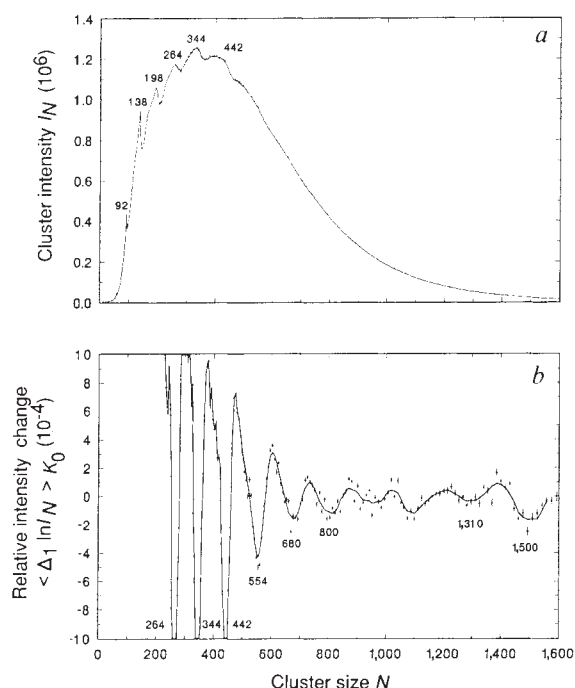


FIG. 1 Mass spectrum and its logarithmic derivative. *a*, Example of an abundance distribution for sodium clusters produced by adiabatic expansion and measured after 1 m free flight by time-of-flight mass spectrometry. *b*, Logarithmic derivative $\langle \Delta_1 \ln I_N \rangle_{K_0}$ of the results in top panel according to eq. (2).

provides additional evidence for the existence of quantum shells up to $N \leq 2,500$ in sodium.

Our experiment is based on measurements of the size distribution of metal droplets produced by expansion of metal vapour through a fine nozzle. Sodium vapour is generated in a stainless steel vessel held at 700–800 °C, and the expansion is assisted by pressurizing the vessel with a large surplus of argon or xenon gas. The mixture is expanded into a 3-m-long, differentially pumped flight tube, resulting in a narrow beam of free-flying metal clusters with velocities close to the initial thermal speed of the argon or xenon atoms⁷. In the expansion and clustering process, lasting about 100 ns, the noble gas medium cools to a few tens of kelvin². In the sodium clusters, on the other hand, there will be a competition between heating, due to condensation, and cooling, due to collisions. We believe that the resulting internal temperature in the freshly formed clusters

is ~100–200 °C below the oven temperature. In the ensuing free flight (for ~1 ms) the droplets will therefore lose sodium atoms by stepwise evaporation, cooling to ~100–200 °C. This evaporation process is sensitive to shell-like variations in the atomic separation energies; and these variations are thought to be responsible for the step-like modifications of the experimentally observed size distributions^{7,10}, although the pre-evaporation size distributions are presumably completely smooth.

The size distribution is sampled 1 m downstream by time-of-flight mass spectrometry⁷. Ultraviolet photons with energies close to the ionization threshold and an energy spread of 1 eV (compared with a total Fermi energy in sodium of 3.24 eV) are used to produce a representative sample, in the form of ions, from the otherwise neutral size distribution. Figure 1*a* shows an example of an abundance distribution I_N against N obtained in this way. For sizes up to $N = 300$, the bell-shaped abundance distribution is clearly scarred with saw-tooth or *s*-shaped irregularities at certain 'magic' sizes. As the interest focuses on these, it is convenient to display the experimental result in terms of relative intensity changes:

$$\Delta_1 \ln I_N = \ln(I_{N+1}/I_N) \approx 2 \frac{(I_{N+1} - I_N)}{(I_{N+1} + I_N)} \quad (1)$$

This makes the magic numbers, N_0 , stand out very clearly as dips⁷. As one sees from Fig. 1*a*, the steps signalling shell closures tend to be obscured by noise due to finite counting statistics for the higher N -values. In this situation we have found it useful to generalize equation (1). This is achieved by calculating a weighted logarithmic derivative for properly spaced mass points N ,

$$\langle \Delta_1 \ln I_N \rangle_{K_0} = \frac{\sum_{K=2K_0/3}^{K_0} 2(I_{N+1+K} - I_{N-K})(2K+1)/(I_{N+1+K} + I_{N-K})}{\sum_{K=2K_0/3}^{K_0} (2K+1)^2} \quad (2)$$

choosing values of $K_0 = 0.03N$. The derivative is thus sampled over intervals of $2K_0 + 1$, which is $\pm 3\%$ of the actual size. In this way we can scale up the measured derivatives in order to display small irregularities in the intensity pattern. Figure 1*b*

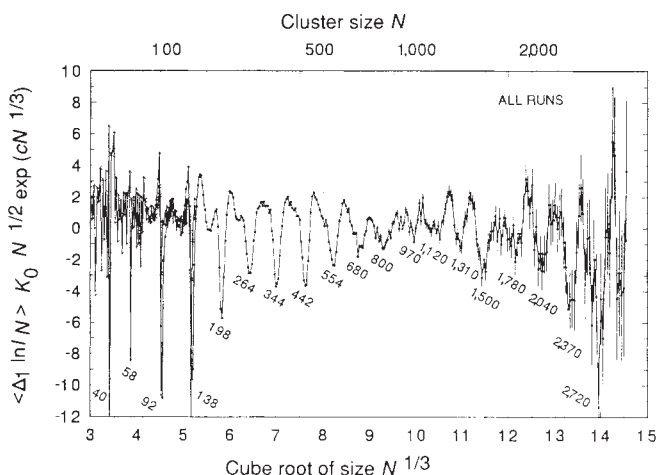


FIG. 2 The quantum beat. Relative changes $\langle \Delta_1 \ln I_N \rangle_{K_0}$ in experimental cluster abundance I_N , corrected for the effect of temperature and size. The model considerations behind the corrections are described in refs 11–13. Measurements from ref. 7 are also shown (crosses). The logarithmic amplitudes $\Delta_1 \ln I_N$ are scaled by an extra factor of 0.2 to account for the superior mass resolution and the use of equation (1) instead of equation (2) for these data.

shows the result. There is a strong decrease in the amplitudes of the shell dips with little or no sign of a beat mode.

There are two reasons for this decrease. First, shell structure is due to periodic variations in an otherwise uniform single-particle electron level density. In a medium of constant density with constant Fermi energy, or, equivalently, constant width of the conduction band, the spacing $\hbar\omega_{\text{shell}}$ between consecutive modulations will decrease as $1/n$ where n is the number of occupied shells. This will cause a decrease proportional to $N_0^{-1/3}$ in the expected dips^{6,7}, as n is proportional to $N_0^{1/3}$. If furthermore, the amplitudes of the shell modulations decrease as $N_0^{-1/6}$, the overall decrease will vary¹¹ as $N_0^{-1/2}$, even at zero absolute temperature. Second, more importantly, temperature^{10,12,13} tends to wash out the observable shell structure. The effect increases exponentially¹¹ in the effective temperature parameter τ ,

$$\tau = k_B T \frac{2\pi^2}{\hbar\omega_{\text{shell}}} \quad (3)$$

with

$$\hbar\omega_{\text{shell}} \approx 2\varepsilon_F/n \propto N^{-1/3} \quad (4)$$

Here, k_B is Boltzmann's constant, T is the real temperature at the time of sampling the abundance distributions, estimated¹⁴ to be 400–500 K, and $\varepsilon_F = 3.24$ eV is the Fermi energy.

To compensate for these effects, we have scaled the experimental logarithmic derivatives, $\langle \Delta_1 \ln I_N \rangle_{K_0}$ (see Fig. 1b) by the factor $N^{1/2} \exp(cN^{1/3})$, setting $c = 0.65$. The results are presented in Fig. 2. They are plotted as a function of the linear dimensions of the clusters, $\sim N^{1/3}$, and comprise all results from our series of experiments, supplemented with some previous results⁷ from the lower size range. The shell dips are equidistant, as expected from theory⁴. In addition, the amplitudes of the dips vary systematically. As the scaling function increases monotonously, it cannot by itself introduce the large-scale modulations seen in Fig. 2. These have the character of a beat mode with a minimum near $N = 1,000$, as predicted by semiclassical theory^{4,6}.

As a further verification of this quantum beat, Fig. 3 presents a plot of the cube root of the magic numbers, $N_0^{1/3}$, against the running index, n , which represents the period (or shell) number. The simplest representation of a beat mode is the sum of two cosine functions of n with slightly different wave numbers k_Δ and $k_\square$, and $k_\Delta + k_\square = 4\pi$:

$$\begin{aligned} \cos(k_\Delta n) + \cos(k_\square n) &= 2 \cos\left(\frac{k_\Delta + k_\square}{2} n\right) \\ &\times \cos\left(\frac{k_\Delta - k_\square}{2} n\right) \end{aligned} \quad (5)$$

The minima of this function will be equidistant, sharing a phase shift of $\frac{1}{2}n$ each time the envelope factor changes sign, that is, with each new superperiod. As seen in Fig. 3, the experimental data show exactly these features. The plot is linear in $N_0^{1/3}$ and there is a phase shift of $1/2n$, which occurs just where the intensities (Fig. 2) pass through a minimum. A plot of the theoretically predicted shell closures⁶ has the same two features. (Note that the results of Fig. 3 can be obtained directly from the experimental data, before any correction for the effect of size and temperature is introduced).

The length of the shell periods, implied by the slope factor (0.61 ± 0.01) from Fig. 3, also supports an interpretation in terms of classical triangular and square orbits. In connection with an earlier study, extending only to $N = 600$, we have shown⁷ that electrons moving classically in triangular and square orbits of total length L will have integer action values, $pL = nh$ for sizes N_s spaced at intervals $\Delta N_s^{1/3} = 0.61$, if they have an average momentum p equal to the Fermi momentum in sodium metal (or a wavelength equal to the Fermi wavelength). For other

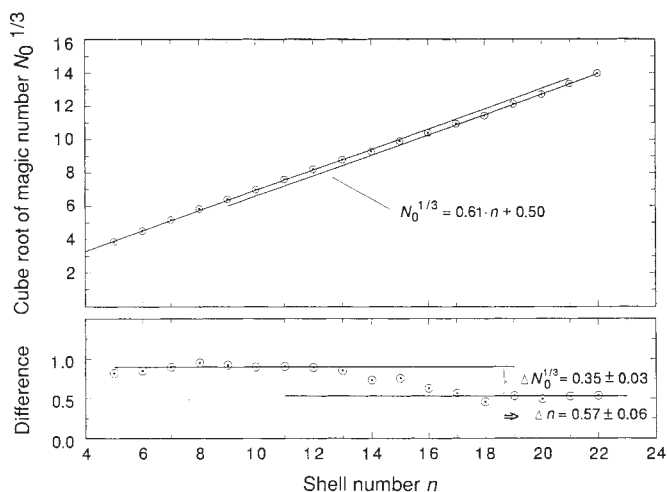


FIG. 3 The phase shift. *a*, Cube root of the magic numbers from Fig. 2, signifying shell closures, plotted against shell number n . For higher shell numbers the points fall on two straight lines, phase-shifted by $\frac{1}{2}n$. This is shown more clearly in *b*, where a straight line has been subtracted from the data above.

sizes, the average momentum and energy of the least bound particles will have to be slightly larger or smaller to fulfill Bohr's quantization condition. This is one way of explaining the shell oscillations. The beat mode comes about because the lengths L of triangular and square orbits differ slightly.

In summary, we have shown that the conduction electrons in droplets of sodium metal form a quantum system with periodic shell structure in analogy to the periodic system of the atomic electrons. The measurements have been extended to drops containing 3,000 conduction electrons. Altogether, 22 shell periods have been identified and found to form a beat pattern in the amplitude of the shell effect, where ~ 10 adjacent shells are grouped together into a supershell. These features agree with the semiclassical theory of Balian and Bloch⁴, describing the quantal shell structure in terms of closed classical orbits. Sodium clusters therefore form a suitable model system for the study of shells with large quantum numbers, where the correspondence between ordered classical motion and quantum shell structure becomes apparent.

The experiments are performed with clusters having finite temperatures, whereas the theories apply to zero temperature. As a result, the observable effects are hundreds of times smaller than the zero-temperature predictions. The survival of measurable effects indicating systematic interference between distinctly different triangular and square orbits despite overwhelming thermal noise is one of the surprising results to emerge from our work. $\square$

Received 9 May; accepted 15 August 1991.

1. Knight, W. D. *et al.* *Phys. Rev. Lett.* **52**, 2141–2144 (1984).
2. de Heer, W. A., Knight, W. D., Chou, M. Y. & Cohen, M. L. *Solid State Phys.* **40**, 93–181 (1987).
3. Kataokuse, I. *et al.* *Int. J. Mass Spectrom. Ion Proc.* **67**, 229–236 (1985).
4. Balian, R. & Bloch, C. *Ann. Phys.* **69**, 76–160 (1972).
5. Bjørnholm, S. *Contemp. Phys.* **31**, 309–324 (1990).
6. Nishioaka, H., Hansen, K. & Mottelson, B. R. *Phys. Rev. B* **42**, 9377–9386 (1990).
7. Bjørnholm, S. *et al.* *Phys. Rev. Lett.* **65**, 1627–1630 (1990).
8. Martin, T. P., Bergmann, T., Göhlich, H. & Lange, T. *Chem. Phys. Lett.* **172**, 209–213 (1990).
9. Martin, T. P. *et al.* *Chem. Phys. Lett.* (in the press).
10. Bjørnholm, S. *et al.* *Z. Phys. D* **19**, 47–50 (1991).
11. Bohr, A. & Mottelson, B. R. *Nuclear Structure Vol. II*, 608 (Benjamin, London, 1975).
12. Brack, M., Genzken, O. & Hansen, K. *Z. Phys. D* **21**, 65–81 (1991).
13. Brack, M., Genzken, O. & Hansen, K. *Z. Phys. D* **19**, 51–53 (1991).
14. Brechignac, C., Cahuzac, Ph., Leygnier, J. & Weiner, J. *J. chem. Phys.* **90**, 1492–1498 (1989).

ACKNOWLEDGEMENTS. We thank O. Echt, W. D. Knight and B. R. Mottelson for advice and encouragement.

Charge migration in supramolecular stacks of peripherally substituted porphyrins

Pieter G. Schouten*, John M. Warman*,
Matthijs P. de Haas*, Marye Anne Fox†
& Horng-Long Pan†

* IRI, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands

† Department of Chemistry, University of Texas, Austin, Texas 78712, USA

PORPHYRIN derivatives play a central part in energy- and electron-transfer processes in natural systems, in which they occur as individual entities or, commonly, as oligomers or supramolecular assemblies. These compounds have also been proposed for use in conducting and photoconducting bulk materials and as conductive and capacitive elements in molecular electronic devices. Much effort has therefore been devoted towards understanding the factors that control energy and charge transport within porphyrin assemblies. Here we describe studies of charge migration along one-dimensional columnar stacks of porphyrin molecules bearing peripheral hydrocarbon groups. In both the solid phase and the relatively plastic liquid-crystalline mesophase, in which the hydrocarbon groups are mobile, charge is transferred between adjacent porphyrin groups with a jump time of a few picoseconds or less. The isotropic liquid phase, on the other hand, is not conductive. The formation of supramolecular structures therefore seems to be necessary to support and direct charge and energy migration in these systems.

The molecular structures of the peripherally octakis- β -nonoxyethyl-substituted, metal-free and zinc porphyrins studied here are shown in Fig. 1; hereafter we will refer to these as H₂P and ZnP respectively. The methods of preparation and characterization of analogous compounds, by differential scanning calorimetry and polarization microscopy, have been reported previously¹. On heating, H₂P undergoes a single phase transition from the crystalline solid to the isotropic liquid at 92 °C. The zinc derivative, however, displays two well separated transitions; one at 102 °C, resulting from a change from the crystalline solid to a discotic mesophase, and a second at 152 °C resulting from complete melting to the isotropic liquid. Possible reasons for the different phase characteristics of the

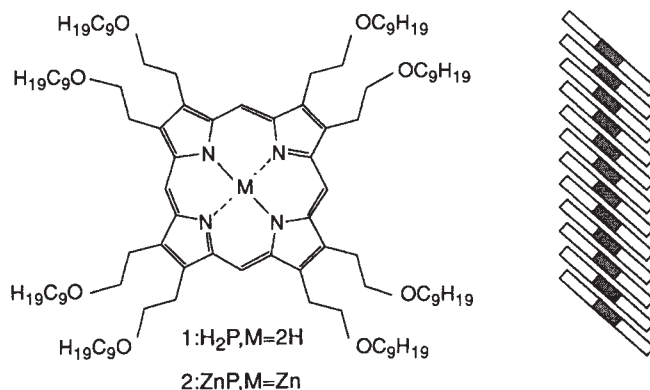


FIG. 1 The molecular structure of the peripherally octakis- β -nonoxyethyl-substituted metal-free and zinc porphyrins used here. Also shown is a schematic representation of the tilted columnar stacking that characterizes the crystalline solid and liquid crystalline phases of these discotic molecules.

H₂P and ZnP compounds have been discussed previously¹ with no clear-cut conclusion being reached.

X-ray diffraction analysis of ZnP shows columnar stacking of the molecules to occur in both the solid and the liquid crystalline phase. The shortest column-to-column distance is determined to be 26.3 Å. A full analysis of similar X-ray data for a series of mesomorphic octa-alkyl and octa-alkoxymethyl derivatives of phthalocyanine² has shown that in both the solid and the mesophase the molecules are tilted at an angle of 46° in the columns, as illustrated in Fig. 1. The tilted conformation results in a centre-to-centre distance between aromatic moieties in the direction of the columnar axis of 4.9 Å.

The present materials in their pure form are insulators with a very low intrinsic conductivity. To study their conductive properties, a small concentration (less than 10 μ M) of electron-hole pairs was produced in the material by a short (2–20-ns) pulse of ionizing radiation (3 MeV electrons from a Van de Graaff accelerator). Any change in the conductivity of the sample resulting from the presence of mobile charge carriers was monitored using the contactless, time-resolved microwave conductivity technique^{3,4}. The frequency band used in the present work was 26.5–38 GHz.

Both the metal-free and zinc porphyrin solids have readily measurable conductivity changes on irradiation. The conductivity decays on a timescale much longer than the pulse width, as is shown in the inset to Fig. 2. The end-of-pulse conductivity per unit energy absorbed, $(\Delta\sigma/D)_0$, is 1.6×10^{-7} and $2.1 \times$

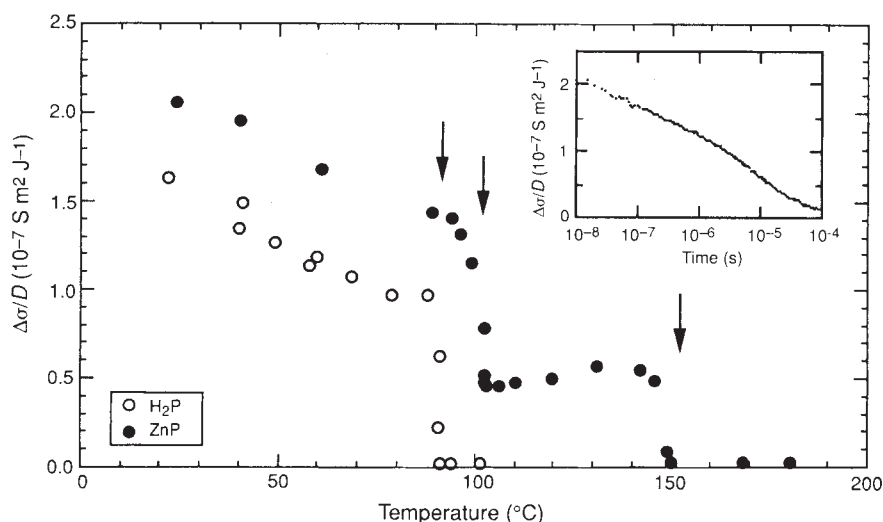


FIG. 2 The temperature dependence of the radiation-induced conductivity of the metal-free (○) and zinc (●) porphyrin compounds shown in Fig. 1. The vertical arrows indicate the temperatures of the phase transitions as determined by differential scanning calorimetry (see text). Inset, the decay of conductivity following the pulse for ZnP at room temperature.

$10^{-7} \text{ S m}^2 \text{ J}^{-1}$ for H_2P and ZnP respectively at room temperature. The presence of the central zinc atom therefore seems to increase the conductive properties of the solid slightly.

In Fig. 2, $(\Delta\sigma/D)_0$ is plotted as a function of temperature. For H_2P , a single, sharp decrease occurs at the transition from the solid to the isotropic liquid. No conductivity could be detected in the isotropic phase. For ZnP , the conductivity decreases abruptly, but does not disappear completely, at the solid to liquid-crystal transition temperature. The conductivity in the mesophase is a factor of ~ 3 less than in the solid with $(\Delta\sigma/D)_0 = 0.5 \times 10^{-7} \text{ S m}^2 \text{ J}^{-1}$. This increases slightly with further increase in temperature in the mesophase and then suddenly drops to zero at the clearing point of ZnP .

We conclude that the long-range columnar order of porphyrin moieties in the solid and liquid-crystalline phases is essential for the conductive properties observed. In the isotropic liquid phase, any molecular organization that might occur must be of such a limited temporal or geometrical extent that it cannot support charge conduction. The results provide definitive evidence, therefore, for the general supposition that electrons and/or holes can move rapidly along the axis of column-stacked π -systems^{5,6}. The decrease in conductivity on going from the crystalline solid to the mesophase is attributed to a lower charge mobility. This is probably due to an increase in positional disorder of the porphyrin moieties⁷ on melting of the hydrocarbon mantle.

We have found a similar decrease in the radiation-induced conductivity at the solid to mesophase transition of octa-alkoxy-substituted phthalocyanines. An isotropic liquid phase of these compounds is not, however, attainable within the temperature range of our equipment. They could not therefore be used to carry out the conclusive experiment of completely melting the material to see if long-range order was essential for conduction. The experiments on octa-*n*-alkoxy substituted phthalocyanines have, however, shown that the lifetime of the conductivity transient increases exponentially with the length of the alkyl tails⁸. This is taken to be strong evidence that the mobile charge carriers responsible are restricted to rapid diffusional motion along the phthalocyanine axis of a stack, with eventual charge recombination requiring stack-to-stack electron tunnelling through the hydrocarbon mantle.

The absolute value of $(\Delta\sigma/D)_0$ is related to the mobilities, μ , of the charge carriers, $\sum \mu = (\mu(-) + \mu(+))$, and the average energy required to produce one charge-carrier pair, E_p (in eV), by^{3,4}

$$(\Delta\sigma/D)_0 = \sum \mu / E_p \quad (1)$$

For organic compounds, the total initial yield of electron-hole pairs corresponds to a value of E_p of $\sim 25 \text{ eV}$ for high-energy radiation⁹⁻¹¹. Only a fraction of the initial electrons and holes will escape rapid (subnanosecond) geminate recombination^{8,9} and diffuse to separate columns, thus being observed in the present experiments. Using a value of $E_p = 25 \text{ eV}$ together with the experimentally determined value of $(\Delta\sigma/D)_0$ will therefore yield a lower limit to the charge-carrier mobilities. For the solid phase of ZnP at room temperature, assuming electrons and holes to have almost equal mobilities¹², this gives $\mu > 2.6 \times 10^{-6} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ and for the mesophase, it gives $\mu > 0.6 \times 10^{-6} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$.

This order of magnitude of the mobility indicates small polaron motion¹³ corresponding to the hopping of a more or less localized charge between neighbouring sites with an average jump time between sites of τ_j . The mobility measured in the present, randomly orientated columnar materials is related to τ_j by

$$\tau_j = ed_j^2 / 6k_B T \mu \quad (2)$$

where d_j is the distance moved along the columnar axis per jump ($4.9 \text{ \AA}$ for the present systems). The minimum values of the carrier mobilities given above correspond therefore to

maximum jump times of 0.6 and 2 ps for solid ZnP at room temperature and the mesophase at 100°C , respectively.

Our results confirm that the transport of charge can occur very rapidly within an organized self-assembly of porphyrin moieties even in a relatively flexible, liquid crystalline phase. Such peripherally hydrocarbon substituted compounds may therefore be considered to provide model systems which crudely mimic the channelling of energy and charge found in the highly varied antenna chlorophyll arrays of photosynthetic membranes. $\square$

Received 9 July; accepted 13 September 1991.

1. Gregg, B. A., Fox, M. A. & Bard, A. J. *J. Am. chem. Soc.*, **111**, 3024-3029 (1989).
2. Weber, P., Guillon, D. & Skoulios, A. *Liq. Cryst.*, **9**, 369-382 (1991).
3. Warman, J. M. & de Haas, M. P. in *Pulse Radiolysis*, Ch. 6. (ed. Tabata, Y.) 101-133 (CRC, Boca Raton, 1991).
4. Warman, J. M. & de Haas, M. P. *Radiat. Phys. Chem.*, **34**, 581-586 (1989).
5. Eley, D. D. *Molec. Cryst. Liq. Cryst.*, **171**, 1-21 (1989).
6. Simon, J. & André, J.-J. *Molecular Semiconductors* (Springer, Berlin, 1985).
7. Pietro, W. J., Marks, T. J. & Ratner, M. A. *J. Am. chem. Soc.*, **107**, 5387-5391 (1985).
8. Warman, J. M., de Haas, M. P., van der Pol, J. F. & Drenth, W. *Chem. Phys. Lett.*, **164**, 581-586 (1989).
9. Warman, J. M. in *The Study of Fast Processes and Transient Species by Electron Pulse Radiolysis* (eds Baxendale, J. H. & Busi, F.) 433-533 (Reidel, Dordrecht, 1982).
10. Schmidt, W. F. & Allen, A. O. *J. Phys. Chem.*, **72**, 3730-3736 (1968).
11. Shinsaka, K. & Freeman, G. R. *Can. J. chem.*, **52**, 3495-3506 (1974).
12. Cox, C. A. & Knight, P. C. *J. Phys. C*, **7**, 146-156 (1974).
13. Mott, N. F. *Conduction in Non-Crystalline Materials*, Ch. 6 (Clarendon, Oxford, 1987).

ACKNOWLEDGEMENTS. We thank M. Northolt and S. J. Picken (AKZO Research Laboratories, Arnhem) for X-ray diffraction analysis of the ZnP compound and for helpful discussions. The synthetic work was sponsored by the Texas Advanced Technology Program.

A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films

Brian O'Regan* & Michael Grätzel†

Institute of Physical Chemistry, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland

THE large-scale use of photovoltaic devices for electricity generation is prohibitively expensive at present: generation from existing commercial devices costs about ten times more than conventional methods¹. Here we describe a photovoltaic cell, created from low-to medium-purity materials through low-cost processes, which exhibits a commercially realistic energy-conversion efficiency. The device is based on a 10- μm -thick, optically transparent film of titanium dioxide particles a few nanometres in size, coated with a monolayer of a charge-transfer dye to sensitize the film for light harvesting. Because of the high surface area of the semiconductor film and the ideal spectral characteristics of the dye, the device harvests a high proportion of the incident solar energy flux (46%) and shows exceptionally high efficiencies for the conversion of incident photons to electrical current (more than 80%). The overall light-to-electric energy conversion yield is 7.1-7.9% in simulated solar light and 12% in diffuse daylight. The large current densities (greater than 12 mA cm^{-2}) and exceptional stability (sustaining at least five million turnovers without decomposition), as well as the low cost, make practical applications feasible.

Solar energy conversion by photoelectrochemical cells has been intensively investigated²⁻¹¹. Dye-sensitized cells differ from the conventional semiconductor devices in that they separate the function of light absorption from charge carrier transport. In the case of *n*-type materials, such as TiO_2 , current is generated when a photon absorbed by a dye molecule gives rise to electron injection into the conduction band of the semiconductor, Fig. 1. To complete the circuit, the dye must be regenerated by

* Present address: Department of Chemistry, University of Washington, Seattle, Washington 98195, USA.

† To whom correspondence should be addressed.

electron transfer from a redox species in solution which is then reduced at the counter electrode. The monochromatic current yield

$$\eta_i(\lambda) = \text{LHE}(\lambda) \times \phi_{\text{inj}} \times \eta_e \quad (1)$$

where LHE (light harvesting efficiency) is the fraction of the incident photons that are absorbed by the dye, ϕ_{inj} is the quantum yield for charge injection and η_e is the efficiency of collecting the injected charge at the back contact, expresses the ratio of measured electric current to the incident photon flux at a given wavelength. The photovoltage ΔV in Fig. 1, generated by the cell, corresponds to the difference between the Fermi level in the semiconductor under illumination and the Nernst potential of the redox couple in the electrolyte.

Although attempts to use dye-sensitized photoelectrochemical cells in energy conversion have been made before, the efficiency of such devices has been extremely low and practical applications have seemed remote. One problem is that of poor light harvesting. On a smooth surface, a monomolecular layer of sensitizer absorbs less than 1% of incident monochromatic light. Attempts to harvest more light by using multilayers of dye have in general been unsuccessful. The remaining option is to increase the roughness of the semiconductor surface so that a larger number of dye molecules can be adsorbed directly to the surface and can simultaneously be in direct contact with the redox electrolyte. Matsumura *et al.*¹² and Alonso *et al.*¹³ have used sintered ZnO electrodes to increase the efficiency of sensitization by rose bengal and related dyes. Willig, Parkinson and colleagues¹⁴ have reported high quantum yields for the dye sensitization of SnS_2 . But the conversion yields from solar light to electricity remained well below 1% for these systems. In addition, the instability of the dyes employed presented a severe practical drawback. By using semiconductor films consisting of nanometer-sized TiO_2 particles, together with newly developed charge-transfer dyes, we have improved the efficiency and stability of the solar cell.

High-surface-area TiO_2 films were deposited on a conducting glass sheet from colloidal solutions. A transmission electron micrograph of the colloid is shown in Fig. 2. Electronic contact between the particles was produced by brief sintering at 450 °C. The size of the particles and pores making up the film is controlled by the size of the particles in the colloidal solution. The internal surface area of the film is determined by the size of the particles and the thickness of the film. These parameters were optimized to obtain efficient light harvesting while maintaining a pore size large enough to allow the redox electrolyte to diffuse

easily. Films of 10 μm thickness consisting of particles with an average size of 15 nm gave linear photocurrent response up to full sunlight and were used throughout. A cubic close packing of 15-nm-sized spheres to a 10- μm -thick layer is expected to produce a 2,000-fold increase in surface area.

Absorption spectra obtained for such nanostructured TiO_2 films are shown in Fig. 3. Bare films are transparent and colourless, displaying the fundamental absorption edge of anatase (band gap 3.2 eV) in the ultraviolet region. Deposition of a monolayer of the trimeric ruthenium complex^{15,16}, $\text{RuL}_2(\mu\text{-(CN)Ru(CN)L'})_2$, **1**, where L is 2,2'-bipyridine-4,4'-dicarboxylic acid and L' is 2,2'-bipyridine, results in deep brownish-red coloration of the film. The absorption onset is shifted to 750 nm, the light harvesting efficiency reading almost 100% in the whole visible region below 550 nm. Integration of the spectral overlap between a solar emission of AM1.5 and this absorption band shows that 46% of the incident solar energy flux is harvested by the dye coated film ($\text{AM} = 1/\sin \alpha$ where α is the angle of incidence of the solar rays at the Earth's surface).

The optical density of the film at 478 nm corrected for the absorption by the conducting glass support was 2.45. Dividing by the extinction coefficient¹⁶ of **1** ($\epsilon_{478} = 1.88 \times 10^4 \text{ cm}^2 \text{ mol}^{-1}$) yields the dye surface concentration, $\Gamma = 1.3 \times 10^{-7} \text{ mol cm}^{-2}$. As each dye molecule occupies an area¹⁶ of 1 nm², the inner surface of the film is 780 cm² for each 1 cm² of geometric surface. Thus, the roughness factor is 780, which is smaller than the predicted value of 2,000. The difference is attributed to necking between TiO_2 particles. In addition, the large size of **1** prevents its access to very small pores, reducing the apparent surface area.

The photocurrent action spectrum obtained with the dye-coated TiO_2 film is also shown in Fig. 3. It closely matches the absorption spectrum, indicating that the current is due to electron injection from **1** into the conduction band of TiO_2 . The photocurrent yield measured at 520 nm was found to depend

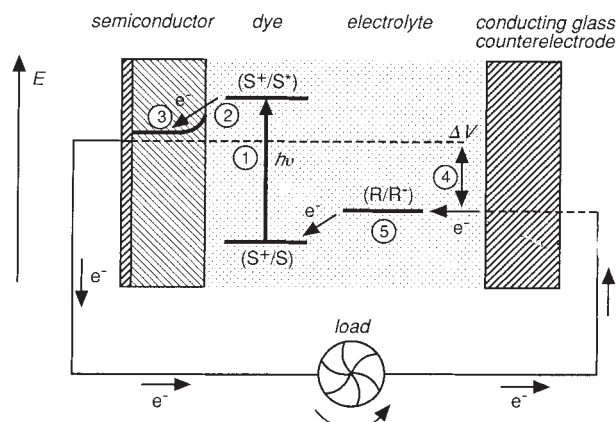


FIG. 1 Schematic representation of the principle of the dye-sensitized photovoltaic cell to indicate the electron energy level in the different phases. The cell voltage observed under illumination corresponds to the difference, ΔV , between the quasi-Fermi level of TiO_2 under illumination and the electrochemical potential of the electrolyte. The latter is equal to the Nernst potential of the redox couple (R/R^+) used to mediate charge transfer between the electrodes. S, sensitizer; S^* , electronically excited sensitizer; S^+ , oxidized sensitizer.

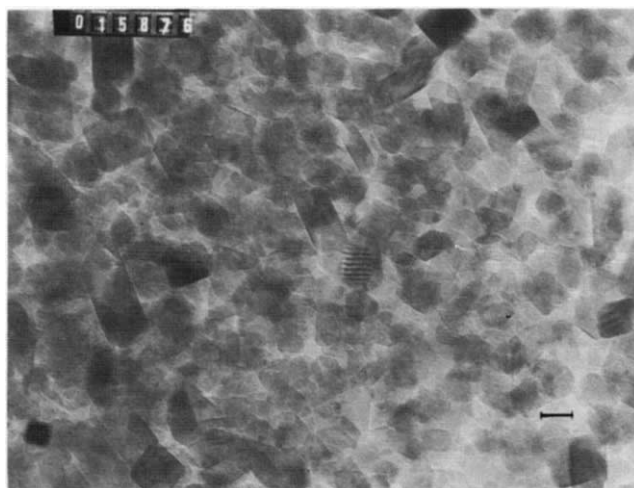


FIG. 2 Transmission electron micrograph of TiO_2 particles used in thin film production. The scale bar represents 10 nm. Particles were prepared by hydrolysis of titanium tetraisopropoxide^{18,19} followed by autoclaving for 12 h at 200 °C. To form films the sol was concentrated by evaporation of water in vacuum at 25 °C until a viscous liquid was obtained. Carbowax M-20,000 (40% by weight of TiO_2) was added and the viscous dispersion (TiO_2 content 20% by weight) was spread on the conducting glass support (Asahi glass, fluorine-doped SnO_2 overlayer, transmission 85% in the visible, sheet resistance 80/square) to give a membrane of 10 μm thickness. This was heated under air for 30 min at 450 °C. High-resolution scanning electron microscopy revealed TiO_2 films to be composed of a three-dimensional network of interconnected nanoscale particles¹⁹. Transmission electron micrographs of TiO_2 particles were taken before heat treatment; the annealing of 450 °C did not induce significant changes in particle size. Before dye coating, a few monolayers of TiO_2 were electrodeposited onto the colloidal TiO_2 film from a Ti(III) solution. Detailed description of this procedure will be published elsewhere (L. Kaven, B.O'R., A. Kay and M.G., manuscript in preparation).

on the counter ion of the iodide/triiodide redox electrolyte, increasing from 68% for tetrapropylammonium to 84% for Li^+ . After correction for the light absorption by the conducting glass, the yields are 80% and 97%, respectively. This shows that the nanostructured TiO_2 films used in conjunction with suitable charge transfer dyes can achieve quantitative conversion of visible light photons into electric current.

Figure 4 shows the current-voltage characteristics obtained with the thin layer cell under illumination by simulated AM1.5 solar light. The conversion efficiencies for one tenth and full sunlight are 7.9% and 7.12%, and the fill factors (maximum output power of cell $\div$ [short circuit current $\times$ open circuit voltage]) are 0.76 and 0.685, respectively. Similar yields were obtained under direct sunlight (measurements performed in June early in the afternoon on the roof of the institute). Under diffuse daylight the efficiency increased to 12%, indicating that under such conditions the cell performance is better than that of conventional silicon devices. This is because the spectral distribution of diffuse daylight overlaps more favourably with the absorption spectrum of the dye-coated TiO_2 film than direct sunlight. The fill factor of the cell remains above 0.7 even at very low light intensity ($<5 \text{ W m}^{-2}$). Conventional photovoltaic cells have a much smaller fill factor (<0.5) under these conditions. This indicates that loss mechanisms such as recombination, normally encountered in semiconductor photoconversion, have been minimized. This result might appear surprising in view of the disordered structure of our film giving rise to defects. But the role of the semiconductor in a dye-sensitized device is merely to conduct the injected majority charge carriers. There are no minority carriers involved in the photoconversion process. Therefore, surface and bulk recombination losses due to lattice defects, encountered in conventional photovoltaic cells, are not observed in such a device.

The long-term stability of cell performance was tested by illuminating the thin TiO_2 film loaded with **1** with visible ($\lambda > 400 \text{ nm}$) light for 2 months. The change in the photocurrent was less than 10% over this period, during which a charge of $62,000 \text{ C cm}^{-2}$ was passed through the device, corresponding to a turnover number of 5×10^6 for the sensitizer. This implies that if any dye degradation had occurred its quantum yield (ϕ_{sec}) is less than 2×10^{-8} . As $\phi_{\text{dec}} = k_{\text{dec}}/\Sigma k$, the rate constant, $k_{\text{dec}} \text{ s}^{-1}$, for excited-state decomposition (due to processes such as ligand loss) must be at least 10^{-8} times smaller than Σk , the sum of rate constants for all channels of dye deactivation. Because charge injection is the predominant channel, this sum is practi-

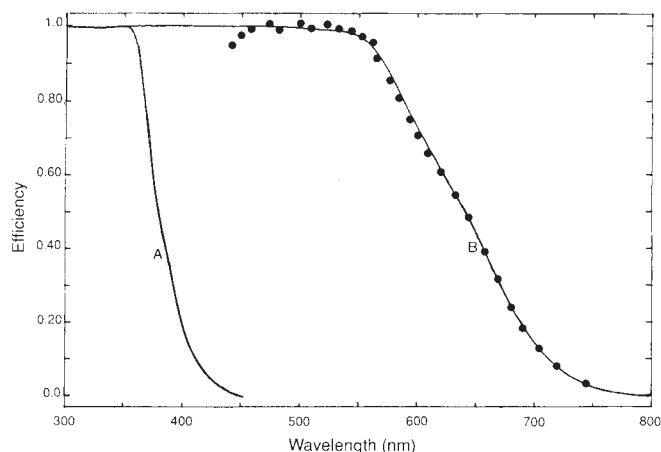


FIG. 3 Absorption and photocurrent action spectra of TiO_2 films supported on conducting glass. A, absorption efficiency of the bare TiO_2 film corrected for conducting glass background; B, absorption efficiency of the same film coated with a monolayer of **1**; full circles, monochromatic current yield at short circuit as a function of excitation wavelength. Yield is corrected for 15% loss of incident photons through light absorption and scattering by the conducting glass support.

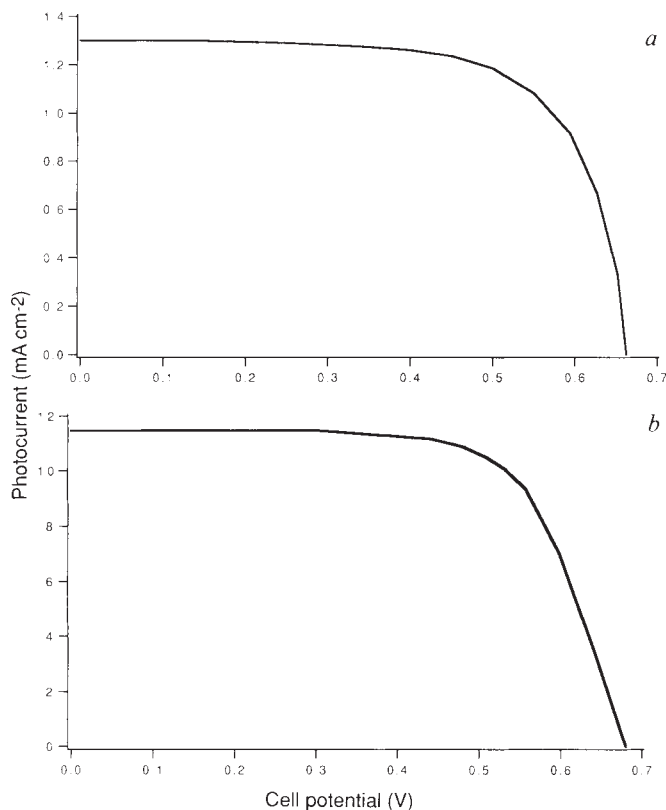


FIG. 4 Photocurrent-voltage characteristics of a cell, based on a colloidal TiO_2 film sensitized by **1**; the film, supported on a conducting glass sheet, was used in a sandwich-type configuration. The size of the dye coated TiO_2 photoanode was 0.5 cm^2 . The counterelectrode, consisting of conducting glass coated with a few monolayers of platinum, was placed directly on top of the working electrode. A thin layer of redox electrolyte is attracted into the intra-electrode space through capillary forces. The cell was exposed to simulated sunlight with AM1.5 spectral distribution. a, light intensity 83 W m^{-2} , electrolyte: 0.5 M tetrapropylammonium iodide + 0.04 M iodine in a mixture of ethylene carbonate (80% by volume) with acetonitrile. Fill factor was 0.76; surface area 0.5 cm^2 (before multiplication by roughness factor). Conversion efficiency was 7.9%. b, light intensity 750 W m^{-2} , electrolyte: 0.5 M tetrapropylammonium iodide, 0.02 M KI $\times$ 0.04 M I_2 in the same solvent. Fill factor was 0.684; conversion was 7.12%.

cally equal to the rate constant for charge injection, which exceeds 10^{12} s^{-1} in the case of **1**. Therefore, the upper limit for k_{dec} is $2 \times 10^4 \text{ s}^{-1}$, which agrees with the known photophysics of this class of transition metal complexes¹⁷. The very fast electron injection observed with dyes such as **1**, combined with high chemical stability, renders these compounds very attractive for practical development. □

Received 19 July; accepted 20 August 1991.

1. Bucher, K. & Fricke, J. *Phys. Zeit.* **21**, 237-244 (1980).
2. Honda, K. & Fujishima, A. *Nature* **238**, 37-39 (1972).
3. Tufts, B. J. et al. *Nature* **326**, 681-683 (1987).
4. Gerischer, H. *Electrochim. Acta* **35**, 1677 (1990).
5. Licht, S., Hodes, G., Tenne, R. & Manassen, J. *Nature* **326**, 863-864 (1987).
6. Heller, A. *Acc. chem. Res.* **14**, 154-162 (1981).
7. Nozik, A. J. *Phil. Trans. R. Soc. Lond.* **A295**, 453-470 (1980).
8. Tributsch, H. & Bennet, J. C. *J. electroanal. Chem.* **81**, 97 (1977).
9. Wrighton, M. S. *Acc. chem. Res.* **12**, 303-310 (1979).
10. Bard, A. J. *Science* **207**, 139 (1980).
11. Memming, R. *Phil. Tech. Rev.* **38**, 160 (1979).
12. Matsumura, M., Nomura, Y. & Tsubomura, H. *Bull. chem. Soc. Japan* **50**, 2533 (1977).
13. Alonso, N., Beley, V. M., Chartier, P. & Ern, V. *Rev. Phys. Appl.* **16**, 5 (1981).
14. Willig, F., Eichberger, R., Sundaresan, N. S. & Parkinson, B. A. *J. Am. chem. Soc.* **112**, 2702-2707 (1990).
15. Amadelli, R., Argazzi, R., Bignozzi, C. A. & Scandola, F. *J. Am. chem. Soc.* **112**, 7099-7103 (1990).
16. Nazeeruddin, M. K., Liska, P., Moser, J., Vlachopoulos, N. & Grätzel, M. *Helv. chim. Acta* **73**, 1788-1803 (1990).
17. Juris, A., Balzani, V., Barigletti, F., Campagna, S., Belzer, B. *Coord. Chem. Rev.* **84**, 85 (1988).

18. Anderson, M. A., Gieselmann, M. J. & Xu, Q. *J. Membrane Sci.* **392**, 43 (1988).
 19. O'Regan, B., Moser, J., Anderson, M. & Grätzel, M. *J. phys. Chem.* **94**, 8720–8726 (1990).

ACKNOWLEDGEMENTS. We are grateful to I. Rodicio for experimental help, to M. Nazeruddin for a sample of sensitizer and to A. Kay for advice on the preparation of colloidal membranes. This work was supported by the Swiss National Energy Office.

Super-spiral structures in an excitable medium

V. Perez-Muñuzuri*, R. Aliev, B. Vasiev,
 V. Perez-Villar* & V. I. Krinsky

Institute of Theoretical and Experimental Biophysics,
 Academy of Sciences USSR, 142292 Pushchino, Moscow Region, USSR
 * Departamento Física de la Materia Condensada, Facultad de Físicas,
 E-15706 University of Santiago de Compostela, Spain

ROTATING spiral waves have been observed in various excitable media, including heart muscle¹, retinae², cultures of the slime mould *Dicystostelium discoideum*^{3,4} and chemical oscillators such as the Belousov–Zhabotinsky (BZ) reaction^{5–7}. Under certain conditions the spiral wave does not exhibit simple periodic rotation, but quasiperiodic⁸ (or ‘compound’⁹) rotation, in which the spiral’s origin (the tip) meanders¹⁰. Recent calculations¹¹ have shown that highly meandering tip motion can impose superstructures on spiral waves. Here we reproduce these patterns experimentally, using the BZ reaction as the excitable medium. We induce high tip meander by applying pulses of electrical current locally at the tip¹². Image processing of the patterns reveals a spiral wave of larger wavelength superimposed on the original wave, an effect that can be described in terms of a Doppler shift in the original spiral.

All our experiments were performed in a BZ reaction¹³ in silica gel at room temperature (20 °C). A 0.9-mm-thick gel layer containing ferroin (0.008 M) was prepared in a Petri dish. During the experiments this layer was covered with a thick (6–8 mm) solution of the other BZ reaction components: 0.1 M NaBrO₃, 0.1 M CH₂(COOH)₂ and 0.5 M H₂SO₄.

The highly meandering tip motion needed for the production of superstructures was enhanced by using stepwise pulses of electric current¹² with a period of 11 min. In each period, the voltage (1.5 V, ~5 mA) was switched on for 4 min.

The electrical circuit used (analogous to that in ref. 12) included two electrodes immersed in the liquid layer. One of these was a circular electrode which ran around the border of the Petri dish, the other was needle-shaped (0.6 mm diameter). The needle was placed over the spiral tip (3 mm into the liquid layer). Two different types of electrodes, stainless steel and platinum, were used, with the same results.

The spiral wave dynamics were followed with a video recorder connected to a PC computer. To observe super-spiral waves, simple image processing was carried out: after acquiring two successive images (separated by 2 min), we subtracted the second from the first. This same procedure was used in the calculations of ref. 11. Contrast enhancement was then used to improve the quality of the printed pictures.

Spiral waves with many turns were created by the following procedure: at the beginning of the experiment, the sulphuric acid concentration was reduced (to ~0.2 M) to avoid the spontaneous appearance of wave sources. Spiral waves were produced in this medium by the breaking of a circular wave. To increase the frequency and the number of spiral turns, sulphuric acid was added until it reached its final concentration (0.5 M). This procedure made it possible to create a single vortex of ~14 turns, with period $T_s \approx 94$ s and wavelength $\lambda_s \approx 0.34$ cm (Fig. 1).

After the medium had become homogeneous, the spiral waves were perturbed by switching the current on and off. This resulted in the successive increasing and diminishing of the size of the core¹², so that the spiral tip described a ‘flowerlet’ pattern. The ratio of two frequencies associated with this motion was about six.

When this highly meandering tip motion occurs, the spiral shape becomes asymmetric. Some of the wavefronts that make up the spiral approach each other (see Fig. 2a). When the picture is processed by subtracting one image from another, a superstructure appears with the shape of a spiral wave (Fig. 2b). This super-spiral does not disappear with time, but rotates as the base spiral does; the tips of both spirals always coincide (Fig. 3). Note that the same results were observed numerically¹¹ (compare upper and lower parts of Fig. 2).

The period and wavelength of the super-spiral ($T_{ss} \approx 540$ s and $\lambda_{ss} \approx 2$ cm) proved to be six times as great as those of the base spiral. In contrast, the linear velocity was roughly equal for both spirals.

The super-spiral can be considered to be the result of the approach between fronts of the base spiral. In this case, the wave period detected at any point in the medium changes periodically with time. Figure 4a shows this effect, measured at a point near the centre of Fig. 2a.

The results presented in Fig. 4a can be interpreted by the Doppler effect. If we suppose that the waves of the spiral are emitted by a source rotating around a circle at constant velocity v_1 and period T_{ss} , then the period of waves at any point in the medium is given by the equation

$$T(t) = T_s \left\{ 1 + \frac{v_1}{v_0} \cos \left(2\pi \frac{t}{T_{ss}} + \phi \right) \right\} \quad (1)$$

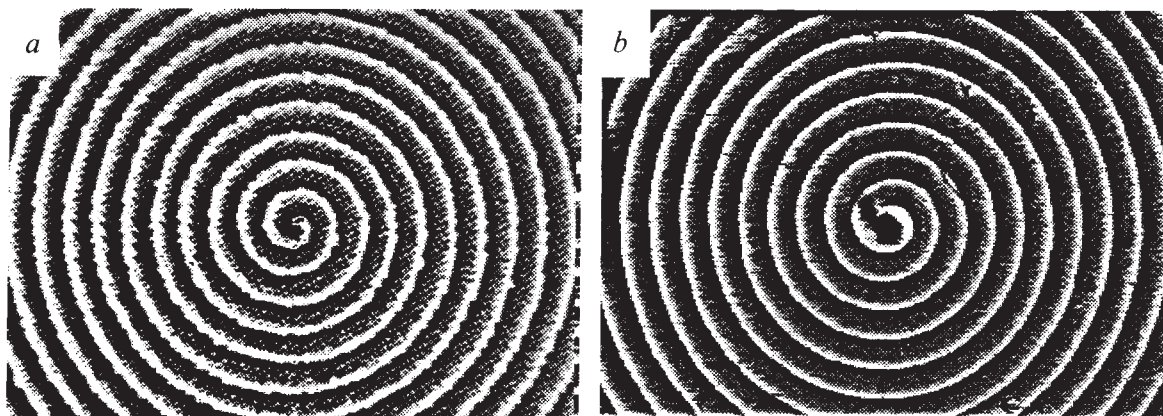


FIG. 1 Rigidly rotating spiral waves (a) for the phase-reaction/diffusion equation, $d\phi/dt = R\{1 - \delta \sin(\phi)\} + \alpha \nabla^2 \phi + \beta |\nabla \phi|^2$ (refs 11 and 15; and P. Hanusse and V.P.M., manuscript in preparation) and (b) in the BZ reaction.

The appearance and behaviour in the two cases are strikingly similar: constant wavelength, sharp fronts and rigid rotation around the centre. The black spot in the centre of b is the electrode shadow.

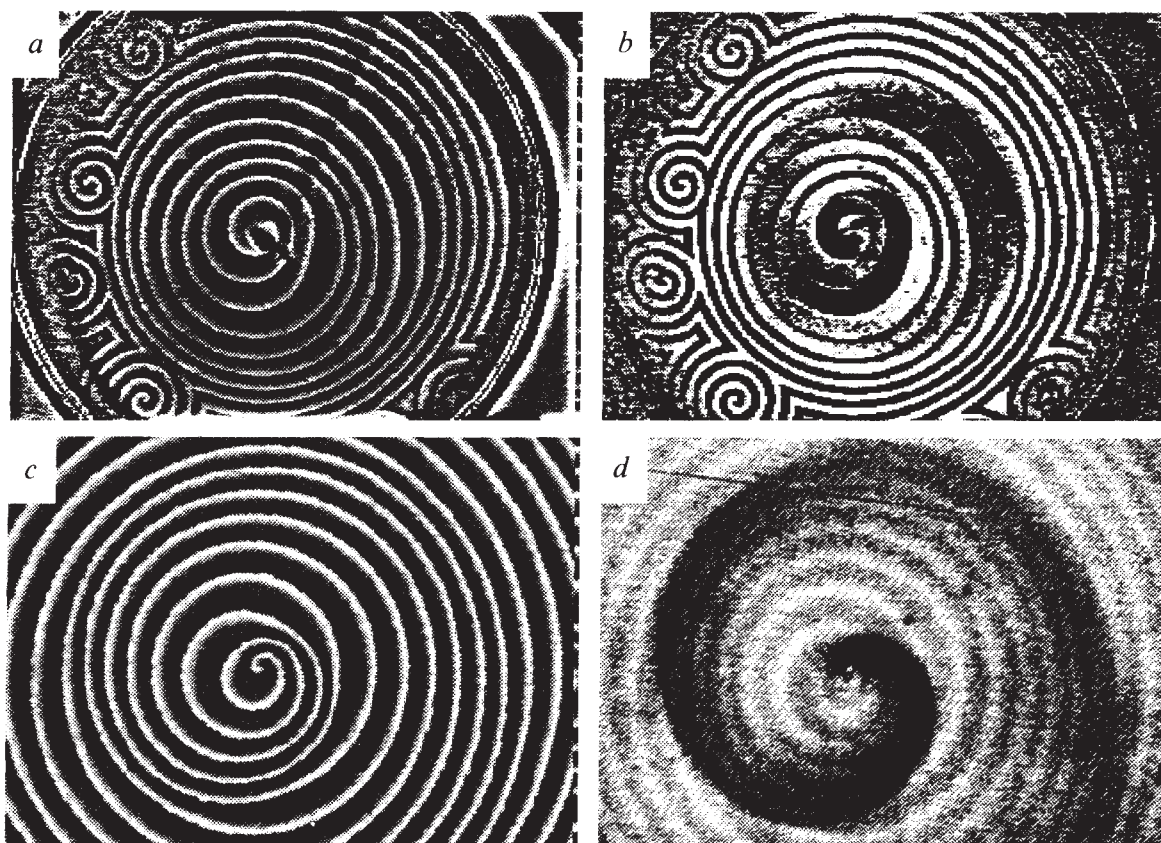


FIG. 2 Experimental and numerical highly meandering spiral and super-spiral waves. *a*, Stepwise electric current (the electrode shadow is seen in the upper part of the picture) was used to enhance the meandering tip motion. *c*, The model¹¹ predicts similar behaviour. Super-spiral waves (*b* and *d*)

superimposed on the non-rigidly rotating vortices (*a* and *c*) can be observed after simple image processing has been carried out (see text). Note that the super-spiral wavelength is six times greater than that of the base spiral.

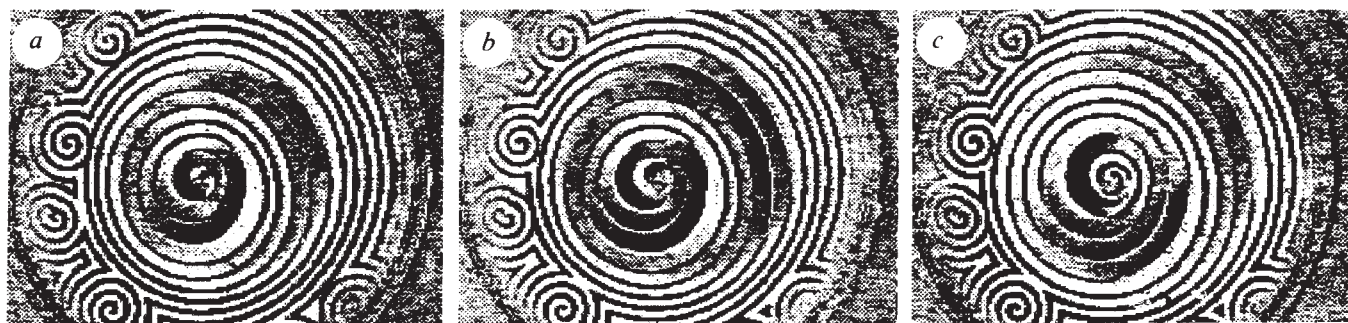


FIG. 3 The rotation of the super-spiral wave in the BZ reaction. The time interval between the frames was 160 s. Far from the core, the super-spiral becomes indistinct because of dispersion and diffusion processes.

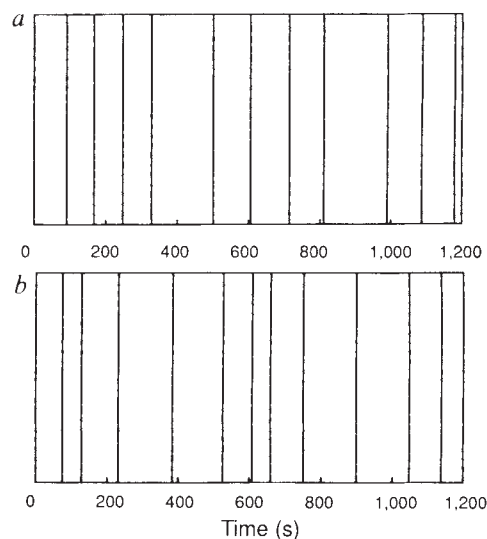


FIG. 4 Doppler effect in BZ reactions. Wavefront positions are represented by vertical lines for (*a*) a point near the spiral centre in Fig. 2*a*, and (*b*) in simulation. In equation (1), $v_1/v_0=0.5$ and $\phi=\pi/3$; other parameters do not affect the observed behaviour.

where v_0 and T_s are the velocity and period of the emitted waves, and ϕ is a phase shift.

Figure 4b shows the periodic behaviour of $T(t)$ when $t = nT_s$ ($n = 1, 2, \dots$). These results are qualitatively similar to those shown in Fig. 4a. In both cases, two regions, with small and large intervals between waves, can be differentiated. The differences between the figures can be attributed to two processes occurring in the BZ reaction: first, the trajectory of the tip is more complicated than in our simulations; and second, diffusion and dispersion effects were not considered.

The effect on the super-spiral behaviour of the relative length of time for which the current is switched on and off has not been studied in detail; work is currently in progress (V.P.M. *et al.*, manuscript in preparation). By controlling this parameter,

it is possible to decrease the wavelength of the super-spiral and observe the transition to spatiotemporal irregular motion, as was described in ref. 11.

Note that electric current was used only to enhance the meandering; its use is not strictly necessary for super-spirals to occur. Such structures can be observed in the BZ reaction even without external stimulation, although in this case the vortices show weak meandering. For such vortices, dispersion and diffusion processes occurring in the BZ reaction result in the waves becoming circular far from the core¹⁴. This causes only the tip of the super-spiral waves to be seen. This restriction may be overcome in other excitable media (such as those listed in the introduction), where the spiral waves exhibit highly meandering tip motion. □

Received 9 July; accepted 29 August 1991.

1. Allesie, M. A., Bonke, F. I. M. & Schopman, T. Y. G. *Circulation Res.* **33**, 54–62 (1973).
2. Goroleva, N. A. & Bures, J. *J. Neurobiol.* **14**, 353–363 (1983).
3. Robertson, A. & Grutsch, A. F. *Cell* **24**, 603–611 (1981).
4. Devreotes, P. N., Potel, M. J. & MacKay, S. A. *Dev. Biol.* **96**, 405–415 (1983).
5. Muller, S. C., Plesser, T. & Hess, B. *Physica D* **24**, 71–86 (1987).
6. Agladze, K. I., Krinsky, V. I. & Pertsov, A. M. *Nature* **308**, 834–835 (1984).
7. Zhabotinsky, A. M., Zaikin, A. N. in *Oscillatory Processes in Biological and Chemical Systems*, Vol. 2 (in Russian) (Nauka, Pushchino, 1971).
8. Zykov, V. S. *Biofizika* **31**, 862–865 (1986); **32**, 337–340 (1987).

9. Skinner, G. S. & Swinney, H. L. *Physica D* **48**, 1–16 (1991).
10. Winfree, A. T. *Science* **175**, 634–636 (1972); **181**, 937–939 (1973).
11. Perez-Muñizuri, V. & Hanusse, P. *J. nonlinear Sci.* (in the press).
12. Perez-Muñizuri, V., Aliev, R., Vasiev, B. & Krinsky, V. I. *Physica D* (submitted).
13. Agladze, K. I., Krinsky, V. I., Panfilov, A. V., Linde, H. & Kunert, L. *Physica D* **39**, 38–42 (1989).
14. Meron, E. *Physica D* **49**, 98–106 (1991).
15. Hanusse, P., Perez-Muñizuri, V. & Vidal, C. in *Nonlinear Wave Processes in Excitable Media*, NATO ASI Series B: Physics Vol. 244 (eds Holden, A. V., Markus, M. & Othmer, H. G.) 501–512 (Plenum, New York, 1991).

ACKNOWLEDGEMENT. V.P.M. thanks P. Hanusse for discussions.

A 13,000-year climate record from western Tibet

F. Gasse*, M. Arnold†, J. C. Fontes*, M. Fort‡, E. Gibert*, A. Huc§, Li Bingyan||, Li Yuanfang||, Liu Qing*, F. Mélières¶, E. Van Campo#, Wang Fubao** & Zhang Qingsong||

* Laboratoire d'Hydrologie et de Géochimie Isotopique, Bâtiment 504, Université Paris-Sud, 91405 Orsay Cedex, France

† Centre des Faibles Radioactivités, CEA-CNRS, BP 1, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

‡ Laboratoire de Géographie Physique, Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 05, France

§ Institut Français du Pétrole, B.P. 311, 92506 Rueil-Malmaison Cedex, France

|| Institute of Geography, Chinese Academy of Science, Beijing 100012, China

¶ Laboratoire de Géologie, Muséum d'Histoire Naturelle, 51 rue de Buffon, 75005 Paris, France

Laboratoire de Géologie du Quaternaire, Faculté des Sciences de Luminy, Case 907, 13288 Marseille Cedex 2, France

** Department of Geography, Nanjing University, Nanjing, China

ALTHOUGH the Tibetan plateau is important in influencing the atmospheric circulation of the Northern Hemisphere^{1–3}, there are only a few continuous palaeoclimate records available, and these are limited to the plateau's northeastern margin^{4–6}. Here we present a 13,000-yr record from Sumxi Co (western Tibet), constructed from both lake-core and shoreline studies, which shows that conditions in the early-middle Holocene were warmer and wetter than at present. These results confirm model predictions of an intensified monsoon over the region at ~9,000 yr BP, owing to an orbitally induced increase in summer insolation^{7,8}. We also find evidence for warm, humid pulses at ~12,500 and ~10,000 yr BP, in phase with the steps of the last deglaciation, and for a return to cold, dry conditions at ~11–10,000 yr BP, none of which can be explained by orbital variations. The existence of the cold episode confirms that the cooling associated with the Younger Dryas event occurred in continental China^{6,9}, and provides further evidence of the global nature of this event¹⁰.

The 1989 Sino-French expedition (Kunlun-Karakorum Programme, CNRS-Academia Sinica) investigated lakes in western

Tibet (mean elevation >5,000 m above sea level), the coldest and driest part of the Tibetan-Qinghai plateau¹¹. Located in the rain shadow of two high mountain ranges, the Kunlun and the Karakorum (Fig. 1a), the region receives rare summer convective precipitation (<50 mm yr⁻¹) and on exceptional occasions, monsoon rainfall (Table 1). Ratios of evaporation to precipitation range between 20 and 50 (ref. 12). Tibetan lakes have experienced large changes in water level after the last glacial maximum (LGM)^{6,12–15}, in response both to warming which induced melting of previously accumulated ice, and to changes in precipitation-evaporation balances, although some lake fluctuations might have tectonic causes.

Sumxi Co (Co means lake; Fig. 1), a lake of tectonic origin¹⁶, is very sensitive to regional climate changes because (1) it lies outside the direct influence of the highest mountains and large ice caps; (2) neotectonics¹⁶ have not significantly modified its hydrology; (3) it is supplied by melt water from small glaciers¹⁷ and local rainfall on a restricted catchment area (Fig. 1b; Table 1). The glaciers lie on granite, whereas most of the watershed

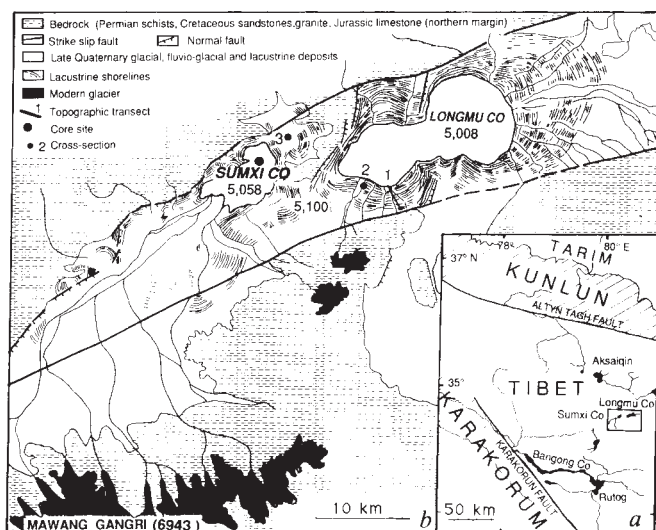


FIG. 1 a, Location map of Sumxi Co in western Tibet. b, The Lungmu Co and Sumxi Co basin.

TABLE 1 Modern characteristics and ^{14}C -dated material from Sumxi Co and Longmu Co

	Modern Lakes			Precipitation	
	Sumxi	Longmu	Runoff Ice meltwater (point 2, Fig. 1b)	Snowfall Longmu Co	Rainfall Rutog (Fig. 1a)
Date	(22/07/89)	(20/07/89)	(21/07/89)	(23/07/89)	(28/07/89)
Altitude (m.a.s.l.)*	5,058	5,008	5,080	5,080	4,243
Lake area (km ²)	24.5	98.7			
Catchment area (km ²)	~2,760	~1,104			
including glaciers (km ²) ¹⁷	143	8.9			
Conductivity (μS cm ⁻¹)	655	150,000	81		
pH	8.7	7.6	8.7		
Dominant cations	Na ⁺ > Mg ²⁺ > Ca ²⁺	Na ⁺ ≈ Mg ²⁺	Ca ²⁺		
Dominant anions	Cl ⁻ ≈ HCO ₃ ⁻	Cl ⁻	HCO ₃ ⁻ + CO ₃ ²⁻		
δ ¹⁸ O _{SMOW}	-3.20	+1.29	-8.85	-12.7	-5.48
δ ² H _{SMOW}	-34.6	-21.1	-54.4	-124.1	-51.6
δ ¹³ C _{PDB}	-0.17				
¹⁴ C (% modern)	91.9 ± 0.6				
AMS ¹⁴ C ages of Sumxi Co and Longmu Co sediments					
SUMXI CO CORE					
Depth (cm)		Material			Age (yr BP)
1,040-1,039		<i>In situ</i> plant remains (Monocotyledon with leaf cuticle)			12,510 ± 190
1,039-1,037		<i>In situ</i> plant remains (mainly <i>Potamogeton pusillus</i>)			12,720 ± 220
791-789		Ostracods (<i>Leucocythere</i>), most shells with connected valves			9,900 ± 420
631-633					6,400 ± 420
373-374		Bivalve (<i>Pisidium</i> sp.) shells, connected valves in pure aragonite			4,770 ± 200
106-108		<i>In situ</i> <i>Potamogeton fluitans</i>			3,400 ± 80
Outcropping sediments					
Altitude (m.a.s.l.)		Material			Age (yr BP)
<i>Shorelines</i> (transect 1, Fig. 1b)					
Bottom calcareous sediments cut off by erosional benches during the lake regression					
5,105		Algal mat			7,290 ± 200
5,098		Ostracods (<i>Limnocythere</i>)			7,520 ± 400
5,080		Algal mat			7,030 ± 220
5,070		Algal mat			7,670 ± 140
<i>Sumxi basin</i> (cross-section 3, Fig. 1b): 4 m of massive calcareous sediments on fluvial deposits					
5,090 base		<i>Potamogeton</i> sp.			8,200 ± 150
middle		<i>Potamogeton</i> sp.			7,260 ± 140
top		<i>Potamogeton</i> sp.			6,890 ± 150
<i>Longmu Co</i> (cross-section 2, Fig. 1b): 2 m of finely laminated calcareous sediments overlain by 3 m of palustral-fluvial deposits					
5,080		Ostracods (<i>Leucocythere</i>)			7,330 ± 470

Summer rain has a monsoon-type content in ^2H and ^{18}O , whereas summer snow shows convective depletion in heavy isotopes^{26,27}. Longmu brine is strongly enriched in heavy isotopes by evaporation. Sumxi is also enriched in ^2H and ^{18}O with respect to melt water, showing the strong influence of evaporation even on fresh waters²⁸. The ^{13}C of the total dissolved inorganic carbon from Sumxi is close to equilibrium with atmospheric CO_2 . Evaporation and exchange at the interface (residence time) provide the main control on the stable isotope content: high ^{18}O and ^{13}C contents of carbonates probably indicate high lake levels. Because there is some carbonate dissolution, the ^{14}C content of the total dissolved inorganic carbon at Sumxi is lower than that of the modern atmosphere. During periods of high residence time, the corresponding dilution effect would have been less than at present, because of enhanced exchange with the atmosphere as shown by the higher ^{13}C content (Fig. 2f). Radiocarbon dates from accelerator mass spectrometry: Tandem, Gif-sur-Yvette. Vegetal remains at 10.39–10.40 m have a thick cuticle of sub-aerial plants. The age consistency with the aquatic *Potamogeton* (10.37–10.39 m) validates datings on material of aquatic origin. The ^{14}C age of the highest sediments, together with their diatom flora with *Cyclotella* sp. 1, confirms the interpretation inferred from the Sumxi Co core.

* Metres above sea level.

consists of schists, sandstones and some carbonates. Sumxi (5,058 m) is a closed freshwater lake, separated from hypersaline Longmu Co (5,008 m) by a threshold at 5,100 m (Fig. 1b; Table 1). Regressive shorelines, visible from 5,140 m, cut off lake-bottom deposits, dated at 7.6–7.0 kyr BP, from 5,105 m downwards (Fig. 1b; Table 1). Sumxi Co and Longmu Co thus formed a single closed lake >132 m deep during the early-middle Holocene.

A 10.53-m piston core, taken in Sumxi Co under 8 m of water, shows an accumulation of laminated sediments overlying indurated palustrine marls (Fig. 2a). Chronological data (Fig. 2a; Table 1), mineralogy (Fig. 2b–d), ^{13}C and ^{18}O contents of authigenic carbonates (Fig. 2e,f), organic matter (Fig. 3a,b), pollen (Fig. 3c), ostracod (Fig. 3d) and diatom (Fig. 3e,f) indicate four main environmental stages.

From ~12.7–12.5 to 10 kyr BP (stage 1), the poor pollen flora is of the type characteristic of deserts (dominated by

Chenopodiaceae and *Ephedra*)^{18–21}. Detrital material predominates, and regular laminae (910–950 cm; Fig. 2a) are made of granite-derived feldspar sand layers, alternating with clays. Seasonal surface discharge of glacier meltwater induced by warming was the main supply to the lake. At ~12.7–12.0 kyr BP, a swamp is recorded by abundant *in situ* plant remains (Table 1) and the presence of littoral ostracods (Fig. 3d)²². Reducing conditions, attributed to long periods of freezing, are indicated by well preserved organic matter (high total organic carbon (TOC) and hydrogen index (HI); Fig. 3a,b)²³ associated with pyrite, and by the lowest ^{13}C content of the section (Fig. 2f). Low ^{18}O contents (Fig. 2e) indicate short residence time causing low evaporation rate^{24,25} and/or a continental effect^{26,27} on a limited amount of atmospheric vapour. Between ~12 and ~11 kyr BP, parallel increases in ^{13}C and ^{18}O contents reflect a longer residence time with evaporation and exchange with the atmosphere²⁸. From ~11 to 10 kyr BP (stage 1B), aquatic

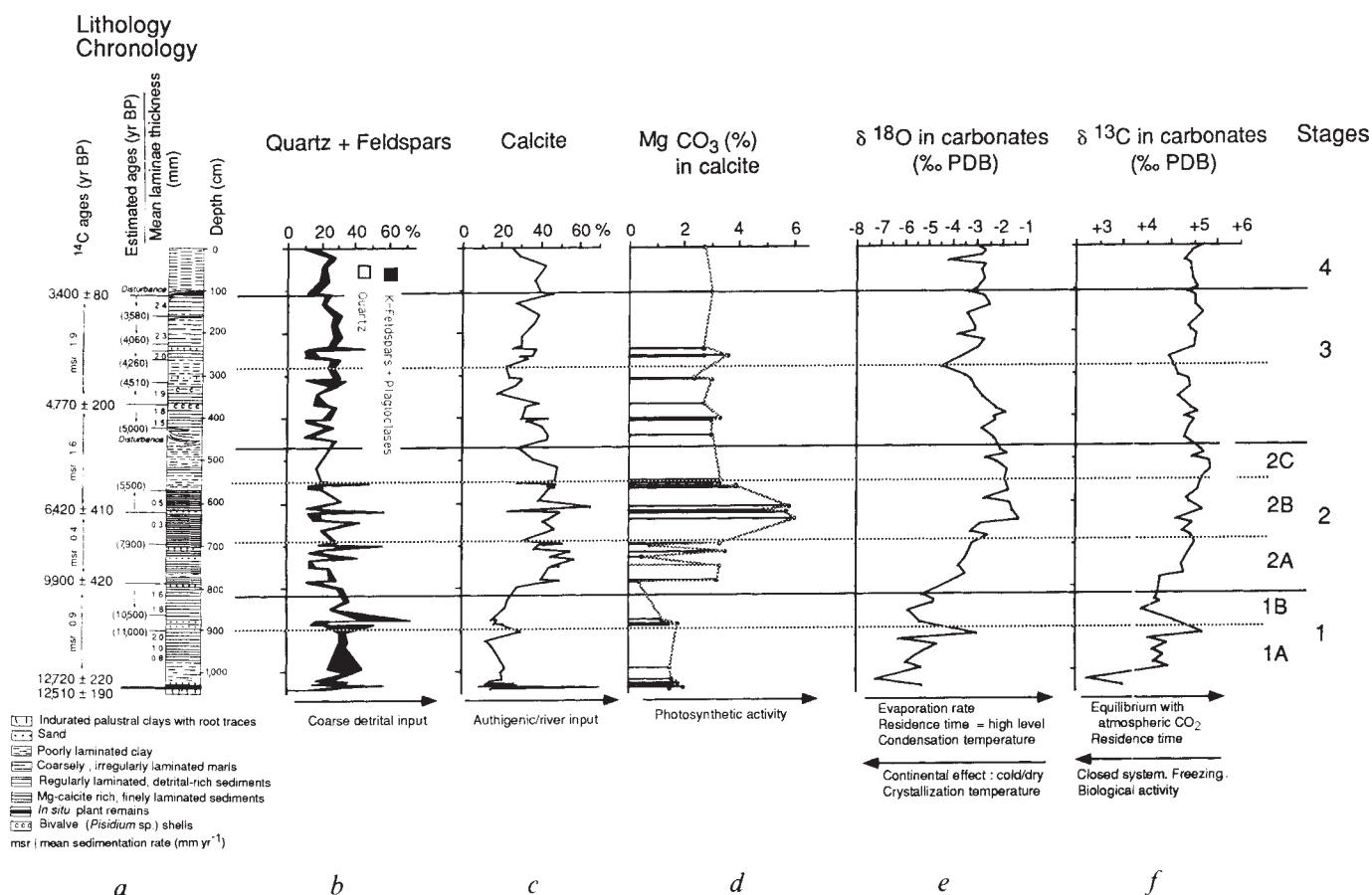


FIG. 2 *a*, Lithology and chronology of the Sumxi Co core. See Table 1 for ^{14}C -dated material. Additional chronostratigraphy is deduced from counting of laminae, their annual nature fitting in with the radiocarbon chronology. *b-d*, Mineralogy (from X-ray diffractometry, using internal standards and peak areas). The calcite Mg^{2+} -content and crystal shape (from scanning

electron microscope observations) indicate its authigeny²⁹. *e, f*, $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ contents of the authigenic carbonate phase (fraction $<40\ \mu\text{m}$). For interpretative arrows, see refs 24–27 and Table 1. $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}}/(^{18}\text{O}/^{16}\text{O})_{\text{PDB}}] - 1$; $\delta^{13}\text{C} = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}}/(^{13}\text{C}/^{12}\text{C})_{\text{PDB}}] - 1$.

organisms and organic matter are absent (Fig. 3*a, d*). The decrease in $\delta^{18}\text{O}$ content (Fig. 2*e*) indicates a cooling, and that in $\delta^{13}\text{C}$ may again reflect the presence of an ice layer during most of the year. Stage 1B is thus attributable to the Younger Dryas cold, dry event.

The early-middle Holocene (~ 10 –5 kyr BP; stage 2) coincides with high water levels (Fig. 1*b*; Table 1). High *Artemisia*/*Chenopodiaceae* ratios (A/C) (Fig. 3*c*) indicate the regional expansion of *Artemisia* steppe^{18–21}, possibly initiated by permafrost decay, but implying an increase in precipitation and/or a change in precipitation regime. Sedimentation rate dropped (Fig. 2*a*) as the lake rose and the herbaceous cover grew, protecting the lake from detrital input. Calcite became the predominant mineral in finely laminated sediments (Fig. 2*a, c*), and intense aquatic biological activity is shown by high ostracod and diatom contents (Fig. 3*d, e*). Diatom development was delayed until ~ 9 –8.5 kyr BP, possibly when a rise in level allowed mixing with the concentrated waters of Longmu Co. From ~ 7.5 to ~ 5.5 kyr BP, diatom abundance, and high values of Mg-calcite (Fig. 2*d*)²⁹, TOC and HI (Fig. 3*a, b*)²³ indicate that primary productivity was optimal. Alternating peaks of littoral ostracod and planktonic diatom contents (Fig. 3*d, e*), however, suggest sudden water-level fluctuation. The highest $\delta^{18}\text{O}$ values of the profile (Fig. 2*e*) reflect high residence time and/or summer-type precipitation from a reservoir of precipitable vapour^{26,27} larger than that of stage 1.

Stage 2 thus corresponds to warm, wet conditions, which were established in two steps: (1) a clear-cut pulse is recorded at ~ 10 kyr BP by sudden increases in A/C ratio, ostracod content, TOC and HI (Fig. 3*a–d*; stage 2A). At ~ 8 kyr BP, decreases in

drier conditions (Fig. 3*c, f*); (2) a new strong increase of the vegetal cover (Fig. 3*c*) and algal productivity (Fig. 3*e, 2d*) occurs at ~ 7.5 –6.0 kyr BP (stage 2B). A reversal of the trends in all environmental variables after ~ 5.5 kyr BP (Fig. 2, 3) is interpreted as a return to drier conditions (stage 2C).

From ~ 5 kyr BP onwards, relatively high TOC with low HI (Fig. 3*a, b*) suggests the weathering of soils²³, and the sediment record may be partly biased by reworking of early-middle Holocene deposits. During stage 3, a diatom assemblage with several species unknown in lower levels indicates a minor water rise at ~ 5 –4.7 kyr BP. Coarsely laminated, clayish-sandy sediments with rare diatom fragments then signify a return to shallow waters. Decreases in the A/C ratio (Fig. 3*c*) and in heavy isotope contents (low residence time) (Fig. 2*e, f*) up to ~ 4.4 kyr BP reflect aridification. Conditions close to present (stage 4) were established after 3.4 kyr BP, with a new diatom community rich in *Fragilaria* sp. found in the modern plankton.

The Sumxi Co data agree reasonably with early-middle Holocene high water levels deduced from southern and northern Tibetan lakes^{6,12–15}, but the concept of maximum lake levels during the Late Glacial in central and western Tibet¹⁴ is not confirmed. Our record is also compatible with a rapid warming at 10 kyr BP and a temperature maximum at 8–6 kyr BP inferred from the Dundee ice cap study⁴.

At Sumxi Co, the early-middle Holocene hydrologic optimum and increase in vegetal cover implies a net increase in the precipitation of evaporation excess, at least at the seasonal scale, attributed to wet and warm summers. The rise in lake water may have been enhanced by ice melting. Increased summer radiation³⁰ and reinforced heat flux from the Tibet land surface

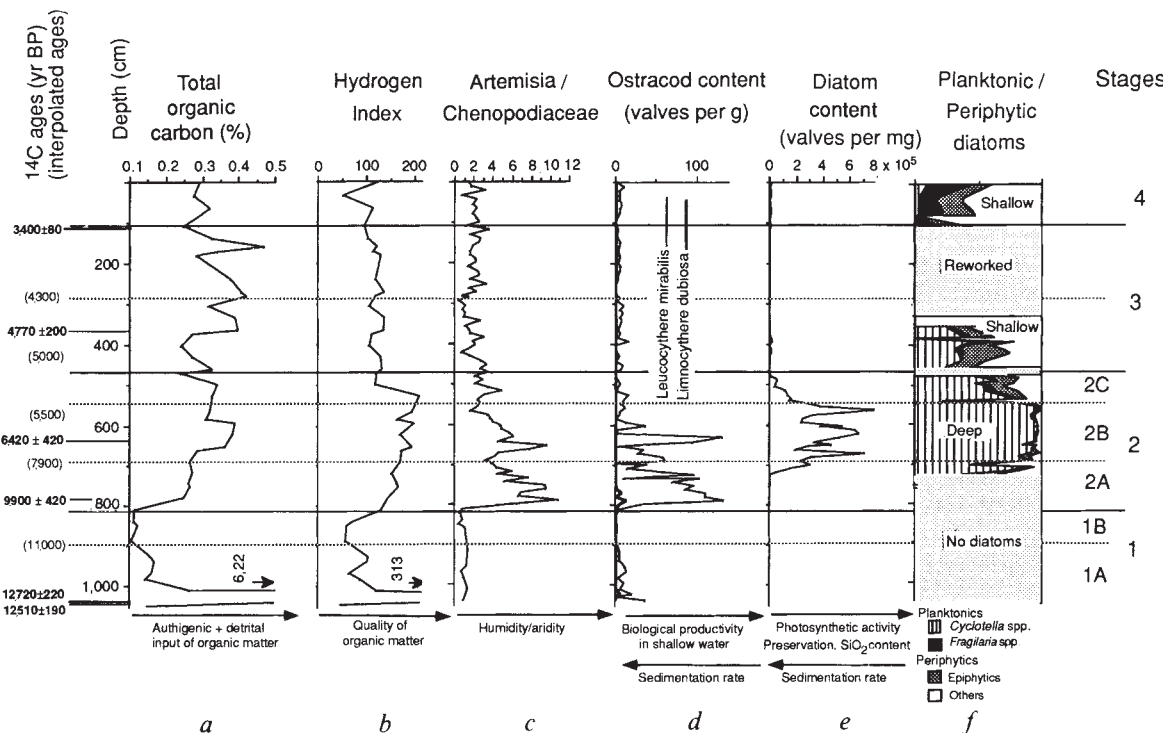


FIG. 3 *a, b*, Total organic carbon (TOC) content, and hydrogen index (HI; mg hydrocarbures per g TOC). Rock-Eval pyrolysis²³. *c*, Pollen percentages: *Artemisia*/Chenopodiaceae ratio. The regional vegetation today comprises desert shrubs (Chenopodiaceae and Compositae dominant), and some alpine plants¹⁸. The two prevalent pollen types of the record, Chenopodiaceae and *Artemisia* (30–90%; mean, 70%) are largely distributed in open formations at low and high altitudes. Chenopodiaceae (C) are characteristic of all types of deserts, and *Artemisia* (A) of steppe communities^{19,20}. Their moisture requirements are different (although both are drought-tolerant, particularly in summer). The ratio (A/C) is thus regarded as an indicator of regional

the A/C ratio and in the planktonic diatom percentage indicate towards today. Condensation may have been enhanced by a recycling of water vapour by lake surfaces and vegetation along the depression pathways. In addition, the season of convective rains would have been longer in response to warmer spring and/or autumn³⁰. Our data thus support the predictions of general circulation models for a regional orbitally induced increase in monsoonal precipitation at 9 kyr BP (refs 7, 8). No clear monsoonal influence is observed at Sumxi Co, however, during the terminal Pleistocene which coincides with the northern summer insolation maximum³⁰. The establishment of the monsoon rainfall on western Tibet would have begun at ~10 kyr BP, lagged by 2–3 kyr behind the precessional forcing.

moisture²¹. *e*, Ostracod content. The two species are littoral in fresh to brackish Tibetan lakes²². *f*, Diatom content. *g*, Diatom percentages. Planktonic forms: *Cyclotella* (percentage of *Cyclotella* sp. 1 reaches 95% during stage 2. This species represents 70–90% of the diatom communities in the modern plankton of lakes Sumxi, Bangong and Aksaiqi (F.G., unpublished data); *Fragilaria* sp. (shallow open water). Periphytic forms: epiphytics (mainly *Cocconeis placentula*, *Amphora pediculus*, *A. libyca*, *Cymbella* sp., *Gomphonema* sp.); epipelics (mainly *Diploneis puella*, *D. ovalis*, *Nitzschia* sp., *Navicula* sp., *Ellerbeckia arenaria*, *Campylodiscus hibernicus* and others).

Superimposed on this general trend, the warm, wet pulses recorded at ~ 12.7 – 12.5 and 10 kyr BP seem to be synchronous with the two main steps of the last deglaciation, deduced from marine cores, and of the arid–humid transition in the northern tropics^{24,31}. These sudden events are thus expected to have global causes, as proposed for the Younger Dryas cold, dry event¹⁰. The presence of this event in continental China, which was suggested from the Qinghai lake⁶ and Loess plateau⁹ records, has been established at Sumxi, in the highest part of the Tibetan plateau. This step-wise climatic evolution possibly requires non-orbital, large-scale controls, such as changes in the atmospheric concentration of greenhouse gases¹⁰, and especially of CH_4 produced in wetlands from endorheic tropical zones. $\square$

Received 30 April; accepted 23 August 1991.

1. Hahn, G. G. & Manabe, S. *J. Atmos. Sci.* **32**, 1514-1541 (1975).
2. Flohn, H. in *Geological and Ecological Studies on Qinghai-Xizang Plateau*, Vol. 2, 1463-1470 (Science Press, Beijing; Gordon & Breach, New York, 1981).
3. Ruddiman, W. & Kutzbach, J. *Scient. Am.* **163**, 46-53 (1991).
4. Thompson, L. G. *et al. Science* **246**, 474-477 (1989).
5. Kelt, K. *et al. Ecol. geol. Helv.* **82**(1), 167-182 (1989).
6. Lister, G. S., Kelt, K., Zao, C. K., Yu, J.-Q. & Niessen, K. *Palaeogeog. Palaeoclimatol. Palaeoecol.* **84**, 141-162 (1991).
7. Kutzbach, J. E. & Otto-Bliessen, B. L. *J. Atmos. Sciences* **39**, 1177-1188 (1982).
8. COHMAP members *Science* **241**, 1043-1052 (1988).
9. Weijian, Z. *et al. Radiocarbon* (abst.) **33**, 259 (1991).
10. Kudrass, H. R., Erlenkeuser, H., Vollbrecht, R. & Weiss, W. *Nature* **349**, 407-408 (1991).
11. Liu Dong-sheng (Ed.) *Geological and Ecological Studies on Qinghai-Xizang Plateau*, Vol. 2, 975-2137 (Science Press, Beijing; Gordon & Breach, New York, 1981).
12. Zheng, M. (Ed.) *Saline Lakes on the Qinghai-Xizang (Tibet) Plateau* (Beijing: Scientific and Technical, Beijing, 1989).
13. Wang, F. B. & Fan, C. Y. *Quat. Res.* **28**, 50-60 (1987).
14. Fang, J.-Q. *Quat. Res.* **36**, 37-60 (1991).
15. Armijo, R., Tapponnier, P. & Han Tonglin *J. geophys. Res.* **94**, 2787-2838 (1989).
16. Liu, Q., Avouac, J. P., Tapponnier, P. & Zhang, Q. *Terra Abstr.* **3** (1), 265 (1991).
17. Li, J. (Ed.) *Glaciers of Xizang (Tibet) (in Chinese)* (Science Press, Beijing, 1986).

18. Wang, C. *The Forests of China, with Survey of Grassland and Desert Vegetation* Maria Moors Cabot Foundation Publ. No. 5 (Cambridge, Massachusetts, 1961).
19. Bottema, S. & Barkoudha, Y. *Pollen Spores* **21**, 427-480 (1979).
20. Peterson, G. M. *Quat. Sci. Rev.* **2**, 281-321 (1983).
21. El-Moslimany, A. P. *Rev. Palaeobot. and Palynol.* **64**, 343-350 (1990).
22. Yang, L., Fan, Y. & Huang, B. *Trans. Oceanol. Limnol.* **1**, 20-27 (1982).
23. Espitalié, J., Deroo, G. & Maquis, F. *Rev. Institut Français Pétrole* **41**, 73-89 (1986).
24. Gasse, F., Tèhet, R., Durand, A., Gilbert, E. & Fontes, J. C. *Nature* **346**, 141-146 (1990).
25. Siegenthaler, U., Eicher, U. & Oeschger, H. *Ann. Glaciol.* **5**, 149-152 (1984).
26. Dansgaard, W. *Tellus* **16**, 436-468 (1964).
27. Yurtsever, Y. & Gat, J. R. in *Stable Isotope Hydrology. Deuterium and Oxygen-18 in the Water Cycle*, 103-142 (International Atomic Energy Agency, Vienna, 1981).
28. Gasse, F. *et al.* *Palaeogeog. Palaeoclimatol. Palaeoecol.* **60**, 1-46 (1987).
29. Bathurst, R. G. C. *Carbonate Sediments and their Diagenesis, Developments in Sedimentology* **12** (Elsevier, Amsterdam, 1975).
30. Berger, A. *J. Atmos. Sci.* **35**, 2362-2367 (1978).
31. Street-Perrott, F. A. & Perrott, R. A. *Nature* **343**, 607-611 (1990).

ACKNOWLEDGEMENTS. We are grateful to Sun Shu and Zheng Du (Academia Sinica), G. Aubert (Institut National des Sciences de l'Univers), P. Tapponnier (Institut de Physique du Globe de Paris) and O. Dolfus (Université Paris 7) for the organization of the 1989-1992 Sino-French Kunlun-Karakorum Programme, Academia Sinica-CNRS. We thank J. F. Dobremetz, J. P. Avouac and L. Bourjot for assistance with field work. CNRS is acknowledged for financial support.

High internal pressures in diamond fluid inclusions determined by infrared absorption

Oded Navon

Institute of Earth Sciences, The Hebrew University, Jerusalem 91904, Israel

FLUID inclusions in mantle-derived minerals provide a rare opportunity to study deep-seated mantle fluids. Among mantle minerals, only diamond possesses sufficient strength to encapsulate and transport to the surface fluids that were trapped at depths below 100 km (ref. 1). Previous studies of the bulk composition² and internal morphology³ of micro-inclusions in cubic and coated diamonds suggested that they contain fluids, but their depth of origin could not be determined. Here I report infrared spectroscopic evidence for residual internal pressures of 1.5–2.1 GPa within these micro-inclusions, higher than any reported previously for fluid inclusions. Extrapolation to mantle temperatures indicates fluid pressures of 4–7 GPa, comparable to those estimated on the basis of mineral equilibria between crystalline inclusions in diamond⁴. These pressures fall within the diamond stability field in the upper mantle, suggesting that the deep-seated fluids, rich in H₂O, CO₂, SiO₂ and K₂O, are probably the mother solutions from which cubic, and possibly eclogitic diamonds grow.

Geobarometers based on mineral equilibria allow calculation of the depth of formation only for diamonds which contain crystalline inclusions of the peridotitic suite. No such barometers are available for diamonds with eclogitic inclusions⁴ or with fluid inclusions. Recently, Liu *et al.*⁵ used Raman spectroscopy to determine the internal pressure for an eclogitic garnet inclusion. The near-atmospheric pressure they found at room temperature allowed for a wide pressure window (4.7–9.3 GPa) at mantle temperatures. Here I use the infrared shift of quartz to determine the pressure in the micro-inclusions at room temperature and to estimate their depth of origin.

Diamonds of cubic morphology and the coats of coated octahedral diamonds rarely contain discrete crystalline inclusions of either the peridotitic or the eclogitic suites. Rather, these translucent diamonds contain myriads of submicrometre fluid inclusions. Recent examination of such micro-inclusions by transmission electron microscopy³ revealed crystals of apatite, carbonates, phyllosilicates and quartz. A previous infrared study of similar diamonds² found the characteristic absorption bands of apatite, carbonates and phyllosilicates, but revealed no quartz bands at 779 and 798 cm⁻¹ (Fig. 1). Here I propose that the two bands at ~784 and 811 cm⁻¹ in the spectra of most micro-inclusion-bearing diamonds (Fig. 1) are due to quartz absorption, and they have been shifted to higher energies as the result of the high internal pressure currently existing in the micro-inclusions.

Under pressure, the various infrared absorption bands of quartz are shifted to higher energies, and they have been calibrated as barometers for work using diamond anvil cells⁶. Figure 2 compares the position of the two bands in 25 diamonds from Zaire and Botswana with an experimental pressure calibration curve⁶. The shift of both bands yields similar pressures. The shift of the band at 798 cm⁻¹ indicates a pressure of 1.5–2.1 GPa (from the y-axis of Fig. 2). The band at 779 cm⁻¹ (x-axis) is less sensitive to pressure and yields 1.1–2.2 GPa. An alternative pressure calibration⁷ of the band at 798 cm⁻¹ yields 1.3–2.1 GPa. These pressures are all consistent with the presence of quartz rather than coesite, as the latter is stable only at pressures $P > 2.18$ GPa at room temperature⁸.

Such high internal pressures should also affect other bands in the spectra of the micro-inclusions. The broader bands at ~475 and 525 cm⁻¹ in the spectra of the diamonds (Fig. 1) are consistent with shifts caused by pressures of ~2 GPa and

1.2 GPa of the 464 and 515 cm⁻¹ bands of quartz, respectively⁶. Alternatively, as absorption in the region 450–550 cm⁻¹ is typical of many silicates, these bands could be produced by other phases. The carbonates in the inclusions are complex Ca–Fe–Mg-carbonates³. As no infrared data exist for complex carbonates at high pressure, it is impossible, at present, to use infrared either for identification of the carbonate phases or for barometry. I am not aware of any high-pressure infrared data for apatite.

Water and CO₂ are also present in the inclusions². At pressures of 1.5–2.1 GPa, the stable form of water is ice VI (ref. 9). The water band, however, resembles the spectrum of liquid water. The broad peak at ~3,420 cm⁻¹ shows no shift towards that of ice even when cooled to 77 K. It is distinct from the spectra of either ice VI, with its peak at ~3,340 cm⁻¹, or ice VII, which shows a much narrower band at <3,410 cm⁻¹ (ref. 10). I conclude that water is present as a supercooled fluid. Freezing is prevented by the high surface-to-volume ratio of the micro-inclusions and the high solute content of the residual fluid¹. A sharp peak appears in the spectra of many micro-inclusion-bearing diamonds at ~2,360 cm⁻¹. It is commonly accompanied by a weaker shoulder at ~2,340 cm⁻¹. Exact positions and intensities are uncertain, as both peaks are masked by intense diamond absorption. The line at 2,360 cm⁻¹ may be due to the presence of dry ice, which absorbs at 2,345 cm⁻¹ at 1 bar but shifts to ~2,360 cm⁻¹ at 2 GPa (ref. 11). Note, however, that gaseous atmospheric CO₂ absorbs at 2,360 and 2,341 cm⁻¹, so that the presence of solid CO₂ in the micro-inclusions is possible but not certain.

The high pressure within the micro-inclusions is, to the best of my knowledge, the highest pressure ever recorded in any fluid inclusion. The preservation of this high internal pressure was probably made possible by the high strength and stiffness of the diamond combined with the small sizes of the inclusions. Larger inclusions would probably lead to decrepitation of the diamond matrix¹². Extrapolating the decrepitation model of Taylor¹² to sub-micrometre-size inclusions, it seems likely that the stress applied by the micro-inclusions (<7 GPa) could be accommodated elastically by the diamond.

Guthrie *et al.*³ found euhedral crystals within the inclusions.

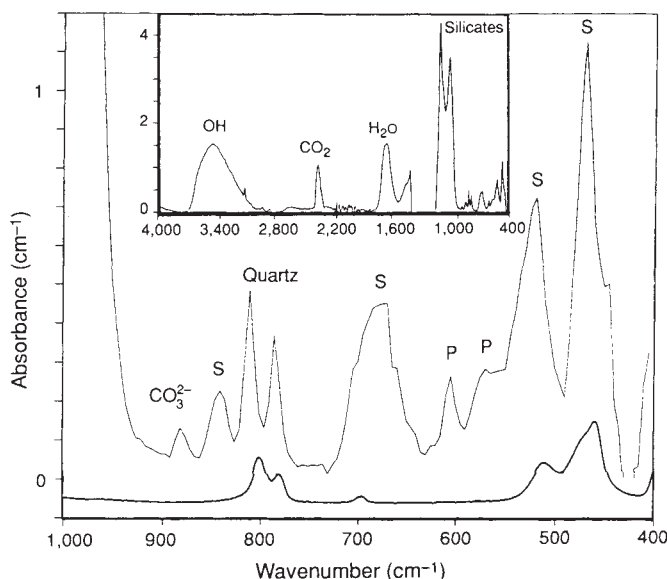


FIG. 1 Infrared spectrum of micro-inclusions in diamond ON-MMZ-83 (thin solid line). The spectrum between 1,000 and 400 cm⁻¹ shows peaks that are related to carbonate (CO₃²⁻), apatite (P), silicates (S) and quartz. The quartz bands are shifted to higher wavenumbers than the bands in a spectrum of pure quartz at 1 bar (bottom spectrum). Inset: the full infrared spectrum of the micro-inclusions with the bands of water and CO₂ marked. (The spectrum of the micro-inclusions is the difference between ON-MMZ-83 and a pure diamond spectrum.)

In no case did the crystals fill the entire volume of the inclusions. These observations and the detection of water and CO₂ by infrared spectroscopy strongly suggest that the pressure within the inclusions is hydrostatic and is exerted by the trapped volatiles. If so, and if the inclusion volume change is known, a pressure-temperature path can be calculated. Calculations show that, to a good approximation, volume is conserved. For example, if diamond is transported from 1,100 °C and 6 GPa to the surface, its volume decreases by 1.4% because of cooling but increases by 1.1% because of decompression (using volume thermal expansion and compressibility of $1.3 \times 10^{-5} \text{ K}^{-1}$ and $1.8 \times 10^{-12} \text{ Pa}$, respectively¹³). In addition, the volume of the inclusions also increases because of the elastic response to the removal of confining pressure. Because of the small size of the inclusions and the strength and stiffness of the diamond, no failure is expected, and to a first approximation $\Delta V/V = 3P_i(1-\nu)/E$, with $\Delta V/V$ being the volume increase, P_i the internal pressure in the inclusions, ν the Poisson ratio and E the Young's modulus. For diamond, $\nu = 0.2$ and $E = 1,050 \text{ GPa}$ (ref. 13). With $P_i = 1.8 \text{ GPa}$, the elastic expansion is 0.4% and the net volume change between mantle and surface is only 0.2%. The effect of this small change on pressure estimates at high temperature is negligible.

As volume is conserved, a pressure-temperature path for the micro-inclusions can be estimated using available equations of state for water-CO₂ mixtures. In the model below, I assume that at room temperature, water is present as metastable liquid water and CO₂ as dry ice. I ignore the effect of the high solute content on water and CO₂ activities, and the effect of the precipitation of the crystalline phases on the volume available for the fluid. I further assume that at high temperature, H₂O and CO₂ do not react to form H₂CO₃. Under these assumptions the isochores for H₂O-CO₂ mixtures are bracketed by the water and the CO₂ isochores.

At 1.8 GPa, 23 °C, a molar volume of $13.9 \text{ cm}^3 \text{ mol}^{-1}$ is estimated from the metastable extension of compression data for

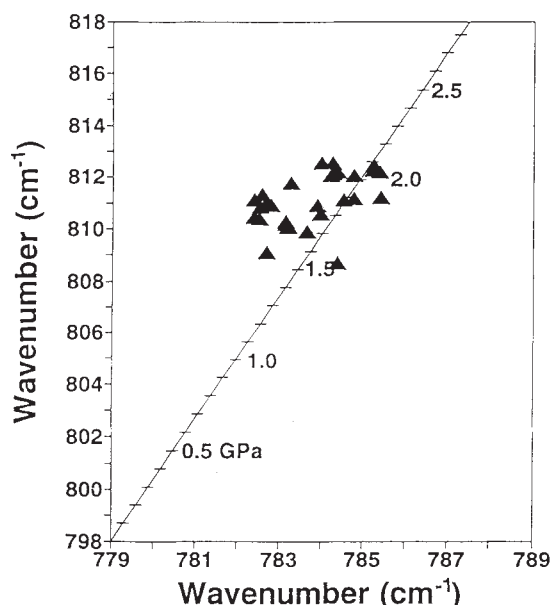


FIG. 2 Infrared shift of quartz absorption bands and internal pressure in the micro-inclusions. The full line, marked with a pressure scale, is an experimental calibration curve for the pressure shift of quartz absorption bands⁶. At 1 bar, the peaks are at 779 and 798 cm⁻¹. With increasing pressure both bands shift along the scaled line. Each horizontal mark represents 0.1 GPa (1 kbar). Triangles show the peak position of the two bands from diamonds. Spectra were recorded on polished diamond wafers (0.4–1 mm thick) on a Nicolet 60SX Fourier transform infrared spectrometer (in Prof. G. R. Rossman laboratory at the California Institute of Technology), and a Nicolet 740 Fourier transform infrared spectrometer in Jerusalem.

liquid water⁹. Extrapolation of high-pressure data¹⁴ for dry ice I yields a molar volume of $24.3 \text{ cm}^3 \text{ mol}^{-1}$, but at 1.8 GPa CO₂ is present as dry ice II (ref. 15) for which P - V data are not available. Metastable extension of volume data for liquid CO₂ yields a volume of $\sim 25.5 \text{ cm}^3 \text{ mol}^{-1}$, and a molar volume of $24.5 \text{ cm}^3 \text{ mol}^{-1}$ is used in the following calculations.

Common equations of state of H₂O and CO₂ fluids that were fitted only to low-pressure data ($< 8 \text{ kbar}$)^{16,17} cannot be extrapolated to such small molar volumes. Figure 3 presents the P - T path along isochores calculated using the equations of Bottinga and Richet¹⁸ (a modified Redlich-Kwong (MRK) equation), Taylor¹² (MRK equations adjusted for high-pressure work), Saxena and Fei¹⁹ (third-order virial equations fitted to shock-wave data at ultra-high pressures), and Belonoshko and Saxena²⁰ (fit of molecular dynamics calculations). From the MRK equation for water¹², pressures in the range of 5.5–6.7 GPa are expected at temperatures of 1,000–1,300 °C. Water-CO₂ mixtures may reach higher pressures, up to 7.5 GPa. Agreement between the two MRK equations^{12,18} for CO₂ is good. The equations of Saxena and Fei¹⁹ yield lower pressures (4.1–4.8 GPa) for both water and CO₂. The more recent estimate of Belonoshko and Saxena²⁰, however, predicts higher pressures (4.4–5.3 GPa). Another molecular dynamics calculation²¹ for a single molar volume of $18 \text{ cm}^3 \text{ mol}^{-1}$ does not allow extrapolation, but predicts even higher pressures than that of ref. 20. In any case, all estimates fall close to the pressure estimates for South African diamonds⁴ (squares in Fig. 3).

The high internal pressure that exists at present in the micro-inclusions indicates that most inclusions are well sealed by the diamond matrix and still preserve the original fluid from which the host diamonds grew. Pressure estimates at reasonable upper-mantle temperatures fall within the diamond stability field. This suggests that the diamonds were not formed metastably in the ascending kimberlites. Rather, the diamonds and the fluid they trapped originate from depths of 130–230 km (4–7 GPa). The finding of a sanidine inclusion in a coated Siberian diamond²² suggests that micro-inclusion-bearing diamonds belong to the eclogitic paragenesis. If so, pressures of 4–7 GPa (at 1,000–1,300 °C) may also be appropriate for the formation of eclogitic diamonds. Strong enrichment of incompatible elements in

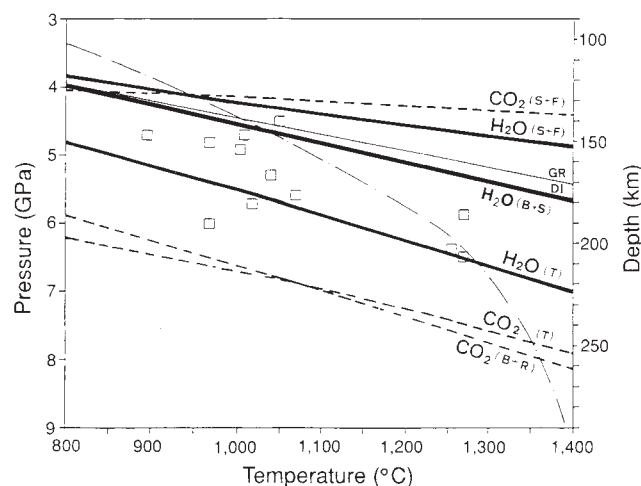


FIG. 3 Possible P - T paths of micro-inclusion-bearing diamonds. Thick solid curves: isochores for water with molar volume of $13.9 \text{ cm}^3 \text{ mol}^{-1}$, using the equations of refs 12 (labelled T), 19 (S+F) and 20 (B+S). Dashed curves: isochores for CO₂ with molar volume of $24.5 \text{ cm}^3 \text{ mol}^{-1}$, according to the equations of refs 12, 18 (B+R) and 19. Water-CO₂ mixtures yield P - T relations that are intermediate between the endmember curves^{12,19}. All curves are either below or just slightly above the graphite-diamond boundary²⁵ (thin solid line). They intersect the 40 mW m^{-2} geotherm²⁶ (dashed-dotted curve) at temperatures of 900–1,300 °C. The defined P - T range for the micro-inclusions, 4–7 GPa and 900–1,300 °C, is similar to that delineated by mineral equilibria of crystalline diamond inclusions (open squares⁴).

eclogitic inclusions in Australian diamonds²³ is also consistent with growth from enriched fluids, similar to those trapped in the micro-inclusions.

The extreme enrichment of incompatible elements in the trapped fluids and their low Mg/Fe ratio² suggest that they were formed as residual fluids in a crystallizing magma. The high pressures reported here place the depth of origin in the diamond stability field and hint at a possible connection with another deep-seated crystallization process: the formation of the megacryst suite. Following the aborted ascent of a highly alkaline

melt, crystallization of olivine, clinopyroxene, ilmenite and other phases leads to extreme enrichment of the residual melt in incompatible elements, water and carbon. Eventually, the melt becomes oversaturated with carbon, and cubic and coated diamonds grow. These late-stage residual fluids are no longer in equilibrium with the wall rock peridotites. If they leave the magma chamber and penetrate the peridotite, they will react with the wall rock and will produce many of the features of metasomatized peridotites. □

Received 26 June; accepted 2 November 1991.

1. Roedder, E. *Fluid Inclusions, Reviews in Mineralogy* Vol. 12 (Mineralogical Society of America, Washington, 1984).
2. Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. *Nature* **335**, 784–789 (1988).
3. Guthrie, G. D., Veblen, D. R., Navon, O. & Rossman, G. R. *Earth planet. Sci. Lett.* (in the press).
4. Meyer, H. O. A. in *Mantle Xenoliths* (ed. Nixon, P. H.) 501–522 (Wiley, London, 1987).
5. Liu, L., Mernagh, T. P. & Jaques, A. L. *Contrib. Mineral. Petrol.* **105**, 156–161 (1990).
6. Velde, B. & Couty, R. *J. non-cryst. Solids* **94**, 238–250 (1987).
7. Wong, P. T. T., Moffatt, D. J. & Baudais, F. L. *Appl. Spectrosc.* **39**, 733–735 (1985).
8. Bohlen, S. R. & Boettcher, A. L. *J. geophys. Res.* **87**, 7073–7078 (1982).
9. Liu, L. *Earth planet. Sci. Lett.* **61**, 359–364 (1982).
10. Wong, P. T. T. & Moffatt, D. J. *Appl. Spectrosc.* **41**, 1070–1072 (1987).
11. Hanson, R. C. & Jones, L. H. *J. chem. Phys.* **75**, 1102–1112 (1981).
12. Taylor, W. R. in *Stable Isotopes and Fluid Processes in Mineralization* (eds Herbert, H. K. & Ho, S. E.) 333–349 (University of Western Australia, 1989).
13. Field, J. E. (ed) *The Properties of Diamond* (Academic, London, 1984).
14. Liu, L. *Earth. planet. Sci. Lett.* **71**, 104–110 (1984).
15. Liu, L. *Nature* **303**, 508–509 (1983).

16. Holloway, J. R. in *Thermodynamics in Geology* (ed. Fraser, D. J.) 161–181 (Reidel, Boston, 1977).
17. Kerrick, D. M. & Jacobs, G. K. *Am. J. Sci.* **281**, 735–767 (1981).
18. Bottinga, Y. & Richet, P. *Am. J. Sci.* **281**, 620–659 (1981).
19. Saxena, S. K. & Fei, Y. *Geochim. cosmochim. Acta* **51**, 783–791 (1987).
20. Belonoshko, A. & Saxena, S. K. *Geochim. cosmochim. Acta* **55**, 381–387 (1991).
21. Brodholt, J. & Wood, B. *Geochim. cosmochim. Acta* **54**, 2611–2616 (1990).
22. Novgorodov, P. G. et al. *Dokl. Acad. Nauk. USSR* **310**(2), 439–445 (1990).
23. Griffin, W. L. et al. *Contrib. Mineral. Petrol.* **99**, 143–158 (1988).
24. Daniels, L. R. M. & Gurney, J. J. in *Kimberlites and Related Rocks*, Vol. 2 (ed. Ross, J.) 1012–1021 (Blackwell, Carlton, 1989).
25. Kennedy, C. S. & Kennedy, G. C. *J. geophys. Res.* **81**, 2467–2470 (1976).
26. Pollack, H. N. & Chapman, D. S. *Tectonophysics* **38**, 279–296 (1977).

ACKNOWLEDGEMENTS. I thank G. R. Rosmann for introducing me to FTIR spectroscopy, for use of his laboratory, and for advice and discussions. J. W. Harris and DeBeers kindly contributed diamonds from Jwaneng, Ph. Chevalier of Sibeka and C. T. Pillinger the diamonds from Zaire. I also acknowledge comments by Z. Reches and E. Stolper, reviews by A. Agnon, I. Gavriel and J. J. Gurney and funding by the Israeli Academy of Sciences and NSF.

Running energetics in the pronghorn antelope

Stan L. Lindstedt^{*†}, James F. Hokanson^{*†},
Dominic J. Wells^{*†}, Steven D. Swain^{*},
Hans Hoppeler[‡] & Vilma Navarro[‡]

^{*} Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming 82071, USA

[‡] Department of Anatomy, University of Berne, CH-3000 Bern 9, Switzerland

THE pronghorn antelope (*Antilocapra americana*) has an alleged top speed of 100 km h⁻¹, second only to the cheetah (*Acionyx jubatus*) among land vertebrates¹, a possible response to predation in the exposed habitat of the North American prairie². Unlike cheetahs, however, pronghorn antelope are distance runners rather than sprinters, and can run 11 km in 10 min, an average speed of 65 km h⁻¹ (ref. 1). We measured maximum oxygen uptake in pronghorn antelope to distinguish between two potential explanations for this ability: either they have evolved a uniquely high muscular efficiency (low cost of transport) or they can supply oxygen to the muscles at unusually high levels. Because the cost of transport (energy per unit distance covered per unit body mass) varies as a predictable function of body mass among terrestrial vertebrates, we can calculate the predicted cost to maintain speeds of 65 and 100 km h⁻¹ in an average 32-kg animal³. The resulting range of predicted values, 3.2–5.1 ml O₂ kg⁻¹ s⁻¹, far surpasses the predicted maximum aerobic capacity⁴ of a 32-kg mammal (1.5 ml O₂ kg⁻¹ s⁻¹). We conclude that their performance is achieved by an extraordinary capacity to consume and process enough oxygen to support a predicted running speed >20 ms⁻¹ (70 km h⁻¹), attained without unique respiratory-system structures.

Maximum oxygen consumption ($\dot{V}_{O_2}^{\max}$) was measured in an open-flow system with the animals galloping up an 11% incline

at about 10 m s⁻¹. Because all animals have an upper limit to the power output that can be fuelled aerobically, when additional power is required beyond the 'maximum aerobic speed', it must be provided anaerobically. This level of power output is recognizable by an increase in lactate production and a plateau in oxygen uptake (this plateau is $\dot{V}_{O_2}^{\max}$). Animals can maintain work outputs of this high intensity only for short intervals.

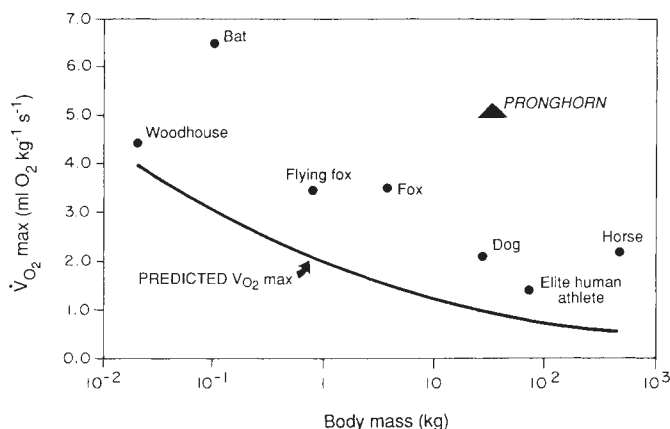
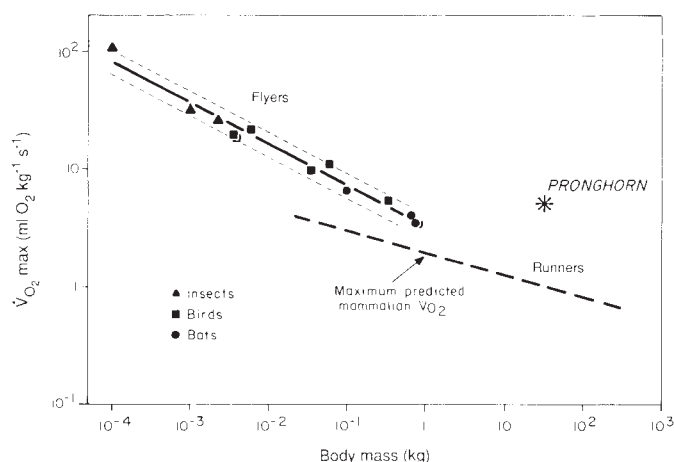


FIG. 1 Maximum weight-specific oxygen consumption ($\dot{V}_{O_2}^{\max}$) varies as a logarithmic function of body mass among mammals; the prediction⁴ is that $\dot{V}_{O_2}^{\max} = 1.92 M_b^{-0.20}$, where M_b is body mass. Because this relationship was derived for a wide variety of mammals, it is not surprising that several notable athletic species, for example dogs, horses and bats^{16–18} fall above the line. $\dot{V}_{O_2}^{\max}$ of two pronghorn antelope (obtained as orphaned fawns from the Wyoming Department of Game and Fish) was determined while each ran on a treadmill while wearing a loose-fitting polyethylene mask. Arterial and mixed venous blood samples were collected (through indwelling catheters) between the third and fourth minute that the animals were running at $\dot{V}_{O_2}^{\max}$ to determine blood gases and oxygen content. From these we calculated cardiac output. Blood lactate was measured immediately after each run to verify that the animals were indeed running at $\dot{V}_{O_2}^{\max}$. Maximum oxygen consumption of pronghorn antelope seems to deviate further from the body-size prediction than that of any other mammal.

[†] Present addresses: Physiology and Functional Morphology Group, Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011, USA (S.L.L. and J.F.H.); Department of Molecular and Cellular Physiology, Royal Veterinary College, University of London, UK (D.J.W.).

FIG. 2 When plotted on log-log coordinates, $\dot{V}_{O_2}^{\max}$ varies as a linear function of body mass among the most highly active (aerobic) invertebrates and vertebrates. Maximum recorded values of oxygen uptake among flying animals seem to fit a single line. The solid line in this figure is created from data taken from wind tunnel measurements (birds and bats) as well as tethered insects, although the latter may well underestimate actual $\dot{V}_{O_2}^{\max}$ (refs 15, 19–26). Despite the difficulties in obtaining true maxima and thus the acknowledgement that this line needs further refinement, the equation describing this relationship, $\dot{V}_{O_2}^{\max} = 3.20 M_0^{-0.35}$ ($r^2 = 0.978$; 95% confidence limits shown), is significantly different [by analysis of covariance, $P < 0.001$] from that given by Taylor *et al.*⁴ (dashed line); the same equation displayed in Fig. 1. Because the extrapolation of the flyers line crosses the runners line at about 30 kg, maximum oxygen uptake of pronghorn antelope is about three times the value predicted by either equation.



The maximum rate of oxygen uptake in these two animals averaged $5.0 \text{ ml O}_2 \text{ kg}^{-1} \text{ s}^{-1}$, over three times the predicted maximum for a similar-sized mammal⁴. Maximum weight-specific aerobic capacity in these animals is about equal to that expected for a 10-g mouse. Though the rate of oxygen uptake is extremely high, locomotor efficiency (or cost of transport) in these animals is apparently unexceptional. The measured rate of oxygen uptake corresponds to that predicted for a mammal to sustain the running speeds observed in this species.

How does the performance of these animals compare with that of other athletic species? To explore this question we consider both body size and mode of transport. Across species, the greatest single determinant of both minimum and maximum rates of oxygen consumption is body size^{4–9}. Thus, it requires over a month for each gram of elephant tissue to consume as much oxygen as does a gram of shrew tissue in a day.

Although body size apparently sets the 'default value' or norm of many structural and functional variables among animals, selective pressures (either natural or man-made by selective breeding) will alter size-dependent constraints. Hence, maximum oxygen consumption in some highly aerobic mammals, such as bats, foxes, dogs and horses, substantially exceed values predicted from body size alone. Deviation from the expected allometric patterns, which Calder¹⁰ refers to as 'adaptive deviation', is extreme in pronghorn antelope (Fig. 1).

In addition to size-dependent patterns in a taxon, $\dot{V}_{O_2}^{\max}$ values also vary as a function of mode of transport across taxa^{11–14}. Just as mammalian runners may be fit to a single line⁴, a single equation can apparently predict the maximum oxygen uptake of flyers, whether insects, birds or bats (Fig. 2). The data included

in this analysis are taken from several references and not all may be true maximum values. Maximum oxygen uptake of pronghorn antelope exceeds that predicted by either runner or flyer regression by threefold.

Is this unusual capacity to consume oxygen the result of some unique respiratory system design? To examine this final question we quantitatively examined structures important in oxygen uptake in pronghorn antelope and a group of similar-sized domestic goats housed in the same facility. Maximum oxygen uptake in the pronghorn was nearly five times that of the goats. Consistent with the observed difference in performance, the pronghorn has proportionately more structure for oxygen diffusion across the lung, greater capacity to deliver the oxygen to the working muscles and more total volume of skeletal muscle mitochondria to utilize the oxygen than do similar-sized goats. Furthermore, the absolute difference in the two species' potential for the uptake, delivery and utilization of oxygen seems to be quantitatively linked to the absolute difference observed in $\dot{V}_{O_2}^{\max}$ (Fig. 3). The one structure that differed the most from the 5:1 ratio was total skeletal muscle mitochondrial volume. This may be explained by a Q_{10} effect due to the much higher muscle (mixed venous blood) temperature in the pronghorn antelope (on average 2.6°C higher than in the goats) as well as potentially tighter packing of inner mitochondrial membranes as is the case in hummingbirds¹⁵. (Q_{10} describes the magnitude of the change in the rate of a biological process for a change of 10°C . Assuming a Q_{10} of 2.5, a 2.6°C difference in muscle temperature would account for a 35% difference in mitochondrial oxygen consumption.) Extreme aerobic capacity in running pronghorn antelope is attributable to extreme design and function of those structures extant in all mammals, it is not the consequence of novelty of design. □

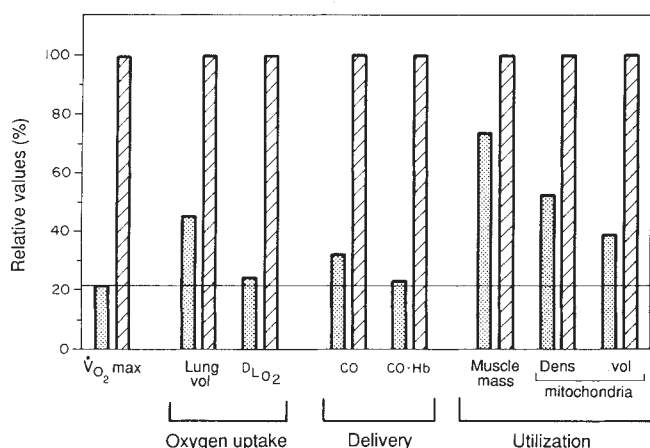


FIG. 3 Pronghorn antelope and domestic goats are similar-sized ruminants that may be closely related² though maximum oxygen uptake is roughly 5 times higher in pronghorn than in a group of goats living at the same facility. We measured key structures critical to supporting $\dot{V}_{O_2}^{\max}$ in both species using techniques perfected and described in detail by Weibel and Taylor²⁷. This histogram presents goat functional and structural data (dotted bars) relative to that of a single pronghorn (hatched bars) (one of the two animals died before the structural sampling). We compare: (1) lung volume and the structural diffusing capacity of the lung for O_2 , (DL_{O_2}) a measure of the lung's ability to transport O_2 from the air into the blood; (2) cardiac output (CO) and haemoglobin concentration (Hb), a measure of the potential delivery of oxygen; and (3) muscle mass and total oxidative capacity (total mitochondrial volume), a measure of the muscles' potential oxygen use. The observed difference we measure in performance ($\dot{V}_{O_2}^{\max}$) is consistent with differences in these structures. Thus it is not necessary to invoke the existence of peculiar structural adaptations to account for the remarkable $\dot{V}_{O_2}^{\max}$ in these animals.

Received 17 June; accepted 27 August 1991.

1. McKean, T. A. & Walker, B. *Resp. Physiol.* **21**, 365–370 (1974).
2. O'Gara, B. W. *Mammal. Species* **90**, 1–7 (1978).
3. Garland, T. Jr *J. Zool.* **199**, 157–170 (1983).
4. Taylor, C. R. et al. *Resp. Physiol.* **44**, 25–37 (1981).
5. Rübner, M. Z. für Biol. **19**, 535–562 (1883).
6. Kleiber, M. *Hilgardia* **6**, 315–353 (1932).
7. Lechner, A. J. *Resp. Physiol.* **34**, 29–44 (1978).
8. Hemmingsen, A. M. *Rep. Steno meml Hosp.* **9**, 7–58 (1950).
9. Hemmingsen, A. M. *Rep. Steno meml Hosp.* **9**, 7–110 (1960).
10. Calder, W. A. III *Size, Function and Life History* (Harvard University Press, 1984).
11. Taylor, C. R., Schmidt-Nielsen, K. & Raab, J. L. *Am. J. Physiol.* **219**, 1104–1107 (1970).
12. Fedak, M. A. & Seeherman, H. J. *Nature* **282**, 713–716 (1979).
13. Tucker, V. A. *Am. Scient.* **63**, 413–419 (1975).
14. Schmidt-Nielsen, K. *Scaling: Why is Animal Size so Important?* (Cambridge University Press, 1984).
15. Wells, D. J. thesis, Univ. Wyoming (1990).
16. Longworth, K. E., Jones, J. H., Taylor, C. R. & Weibel, E. R. *Resp. Physiol.* **77**, 263–276 (1989).
17. Costill, D. L., Fink, W. J. & Pollock, M. L. *Med. Sci. Sport* **8**, 96–100 (1976).
18. Lindstedt, S. L., Wells, D. J., Jones, J. H., Hoppeler, H. & Thronson, H. A. Jr *Int. J. Sports Med.* **9**, 210–227 (1988).
19. Carpenter, R. E. J. *exp. Biol.* **114**, 619–647 (1985).
20. Thomas, S. P. J. *exp. Biol.* **63**, 273–293 (1975).
21. Tucker, V. A. J. *exp. Biol.* **58**, 689–709 (1973).
22. Lasiewski, R. C. *Physiol. Zool.* **36**, 122–140 (1963).
23. Epting, R. J. *Physiol. Zool.* **53**, 347–357 (1980).
24. Casey, M. in *Insect Flight* (eds Goldsworthy, T. Y., Casey, T. M., Goldsworthy, G. J. & Wheeler, C. H.) 257–272 (CRC Press, Boca Raton, Florida, 1989).
25. Rothe, H. J., Biesel, W. & Nachtigall, W. J. *comp. Physiol.* **157**, 99–109 (1987).
26. Tucker, V. A. *Science* **154**, 150–153 (1966).
27. Weibel, E. R. & Taylor, C. R. *Resp. Physiol.* **44**, 1–164 (1981).

ACKNOWLEDGEMENTS. We thank E. R. Weibel and D. Turner, from the University of Berne, Switzerland, for collaboration. Support was provided by NSF, NIH, an American Heart Association Flinn Foundation fellowship and by the Swiss NSF.

Effect of *Steel* factor and leukaemia inhibitory factor on murine primordial germ cells in culture

Yasuhisa Matsui*, Deniz Toksoz†, Satomi Nishikawa‡, Shin-ichi Nishikawa‡, David Williams†, Krisztina Zsebo§ & Brigid L. M. Hogan*||

* Department of Cell Biology, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, USA

† Howard Hughes Medical Institute, Children's Hospital and Dana Faber Cancer Institute, Boston, Massachusetts 02115, USA

‡ Department of Pathology, The Institute of Medical Immunology,

Kumamoto University Medical School, Kumamoto 860, Japan

§ AMGEN Inc. AMGEN Center, Thousand Oaks, California 91320, USA

DESPITE the importance of germ cells to the survival of species, surprisingly little is known about their embryological origin, proliferation, migration and entry into mitotic arrest or meiosis^{1–5}. Mutations in the murine *Dominant White Spotting (W)* and *Steel* genes, which respectively encode the *c-kit* tyrosine kinase receptor and the *c-kit* ligand (or *Steel* factor), impair the development of primordial germ cells (PGCs) *in vivo*, as well as haematopoietic stem cells and neural crest-derived melanoblasts^{3,6–16}. Here we use a monoclonal antibody against *c-kit* tyrosine kinase receptor and recombinant *Steel* factor to study the *c-kit* receptor–ligand system in cultured PGCs. In addition, we show that leukaemia inhibitory factor (also known as differentiation inhibitory activity)^{17,18}, a factor secreted by STO fibroblasts, can stimulate proliferation of primordial germ cells *in vitro*.

We first investigated the effect of soluble recombinant rat *Steel* factor (rSF) on PGCs from 11.5-day-post-coitum (p.c.) mouse embryo gonads. Even after 6 h, the number of adherent alkaline phosphatase-positive cells is significantly higher in wells containing rSF than in controls, showing an effect on PGC attachment (Fig. 1a). The number of PGCs continues to increase in the presence of rSF for 24 h but then declines. The increase

is dose-dependent (Fig. 2), with an optimum at $\sim 30 \text{ ng ml}^{-1}$. Evidence that rSF can act independently of cell attachment is shown in the experiment in Fig. 1c, in which non-adherent PGCs were removed before adding rSF.

In contrast to these results, PGCs from 12.5-day-p.c. embryos do not respond to rSF: the total number decreases over 3 days in both control and treated cultures (Fig. 1b). This suggests that PGCs may lose responsiveness to *Steel* factor at around 13 days p.c. *in vivo*, which is about the time when female and male germ cells in the embryonic gonad enter meiosis or mitotic arrest, respectively.

The effect of rSF was also investigated using younger embryos in which the PGCs are either in the caudal region and the base of the allantois (8.5 days p.c.), or in the hind gut and mesentery (10.5 days p.c.). PGCs in these migratory sites express *c-kit* RNA^{19,20} while the surrounding somatic cells express both *c-kit* and *Steel* factor RNA^{20,21} and genital ridge somatic cells express only *Steel* factor RNA^{20,21}. Because pregonadal PGCs do not grow well on gelatin-coated dishes, we plated them out on feeder layers of non-mitotic mouse STO fibroblasts^{4,5}. Under these conditions, the number of 8.5- and 10.5-days-p.c. PGCs increases for four and two days, respectively, and then declines (Fig. 1d, and data not shown). Others have made similar observations that^{4,5} are consistent with a programmed decline in proliferative capacity of PGCs at around 13 days p.c. *in vivo*. As shown in Fig. 1d, rSF has little or no effect on PGCs cultured on STO cells. This might be a result of the production of *Steel* factor by the STO feeder cells themselves, making exogenous *Steel* factor redundant. In fact, STO cells do express a 6.2-kilobase SCF messenger RNA (data not shown, and ref. 22), which supports this idea.

If STO cells are producing *Steel* factor active on PGCs, then the increase in the number of PGCs on feeder layers should be blocked by the monoclonal antibody ACK2, which recognizes the extracellular domain of the *c-kit* tyrosine kinase receptor and inhibits melanocyte proliferation *in vivo*²³, and haematopoiesis *in vivo* and *in vitro*^{23,24}. As shown in Fig. 1d, this is what happens with 8.5-day-p.c. PGCs on STO cells. Similar results were obtained with 10.5-day-p.c. PGCs on STO cells (data not shown) and 11.5-day-p.c. PGCs cultured without STO cells (Fig. 1c). But the antibody has no effect on either STO cells or somatic cells from the 11.5-day-p.c. genital ridge (data not shown), which is what we expected because *c-kit* RNA is not expressed by these cells *in vivo*^{19,20,22}.

Because STO cells produce *Steel* factor active on PGCs, as an alternative feeder layer we used SI/SI⁴ cells derived from a homozygous *Sl/Sl* null mutant embryo in which the *Steel* (*Sl*) gene is deleted. We also used a derivative of these cells (SI⁴-m220) that expresses only the alternatively spliced form of mouse *Steel* factor in which the extracellular proteolytic cleavage site encoded by exon 6 is missing²⁵. This protein would be associated exclusively with the cell membrane and would not generate soluble *Steel* factor. These experiments lead to the following conclusions (Fig. 3a): (1) SI/SI⁴ cells do not support an increase in PGC number; (2) after 3 days in culture in the presence of saturating amounts of soluble rSF, the number of PGCs is higher on SI⁴-m220 cells than on SI cells, suggesting that membrane-associated *Steel* factor is required in addition to soluble *Steel* factor for optimal effects on PGCs in culture. Finally, STO cells produce other factor(s) affecting PGCs, as STO-conditioned medium increases PGC number even in the presence of excess rSF. STO cells produce leukaemia inhibitory factor (LIF)¹⁷, and the line used here had also been stably transfected with the human LIF gene. We therefore added recombinant human LIF to PGCs cultured on SI/SI⁴ cells and observed an increase in cell number that was synergistic with the increase produced by rSF (Fig. 3b).

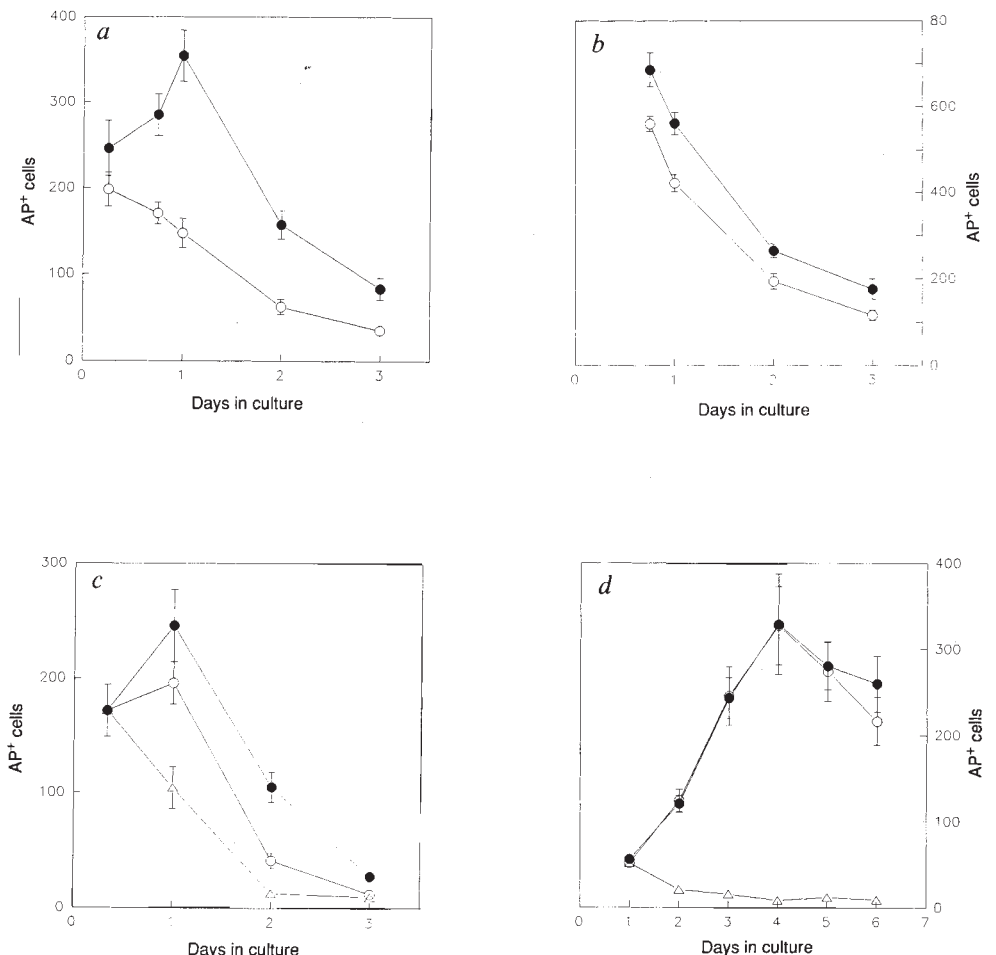
To test whether rSF and LIF act as mitogens or cell survival factors under our culture conditions, which include 15% serum, 8.5-day-p.c. PGCs on SI/SI⁴ cells were incubated for 1 h with

|| To whom correspondence should be addressed.

FIG. 1 The effect of soluble recombinant Steel factor and ACK2 antibody against *c-kit* tyrosine kinase receptor on murine PGCs in culture. Single-cell suspensions of PGCs and somatic cells were prepared from 11.5-day (a, c), 12.5-day (b), or 8.5-day (d) p.c. embryos, and either cultured alone (a, b, c) or on a layer of STO feeder fibroblasts (d). 31.2 ng ml⁻¹ soluble rSF was added at the start of the experiment and the medium changed every 24 h, except in c, when the medium (without rSF) and non-adherent cells were removed after 8 h and replaced with fresh medium with or without rSF or ACK2. ACK2 was added at either 8 h (c) or 24 h (d) to give a final concentration of 1 µg ml⁻¹ IgG2bk. Cultures were fixed and the number of alkaline phosphatase-positive (AP⁺) cells determined⁵. (○) Cultures without rSF, (●) cultures with rSF, and (△) cultures with ACK2. Each experiment was carried out with duplicate wells, and numbers are the means ± s.e.m. of a, 5; b, 5; c, 2; and d, 4; separate experiments.

METHODS. Embryos were from ICR females mated with (C57BL × DBA)F₁ males. Noon on the day of plug is 0.5 days p.c. Genital ridges (11.5–12.5 days p.c.), hind gut and mesentery (10.5 days p.c.) and caudal region (8.5 days p.c.) were dissociated into single cells by incubation at 37 °C with 0.05% trypsin, 0.02% EDTA in Ca²⁺/Mg²⁺-free Dulbecco's phosphate-buffered saline. Cells from the equivalent of 0.5 embryos were seeded into gelatin-coated wells of a 24-well plate (Falcon), containing 1 ml of Dulbecco's modified Eagle's medium, 2 mM L-glutamine, 1 mM sodium pyruvate, 100 i.u. per ml penicillin, 100 µg ml⁻¹ streptomycin and 15% fetal bovine serum. In some cases wells already contained 1 × 10⁵ irradiated or mitomycin C-treated

mouse STO fibroblasts transfected with the bacterial *neo^r* gene and human LIF (provided by A. Bradley). rSF (rrSCF¹⁶⁴) was produced in *Escherichia coli*⁸ and purified. Antibody ACK2 was prepared as described^{23,24}.



bromodeoxyuridine (BrdU) and the number of alkaline phosphatase-positive cells that had incorporated BrdU were counted. There was a significant increase in the proportion of PGCs in S phase in cultures treated with rSF and/or LIF compared with controls (Fig. 3c). We conclude that under certain conditions soluble Steel factor and LIF can act as mitogens for PGCs.

Taken together, our results show that the *c-kit*-encoded receptor/Steel factor ligand system is necessary for the proliferation of PGCs *in vitro*, which accounts for the deficiency of germ cells

in embryos homozygous for severe alleles of *W* and *Sl*. We also show that LIF can act synergistically with soluble Steel factor on PGCs in culture. This is consistent with results on the effect of rSF on haematopoietic stem cells, which demonstrate synergism with other cytokines^{10,16}. But the PGC proliferation *in vitro* is still only a fraction of that *in vivo*, where the numbers increase from about 8 to 25,000 between 7 and 13.5 days p.c.¹⁻³. How further PGC growth is achieved and what factors are involved in survival are matters of great interest.

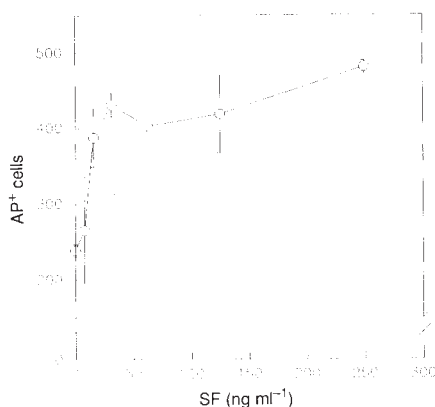


FIG. 2 Effect of different concentrations of rSF on numbers of PGCs. Conditions were as described for Fig. 1a, except that different amounts of rSF were added and the number of alkaline phosphatase-positive (AP⁺) cells was determined after 24 h.

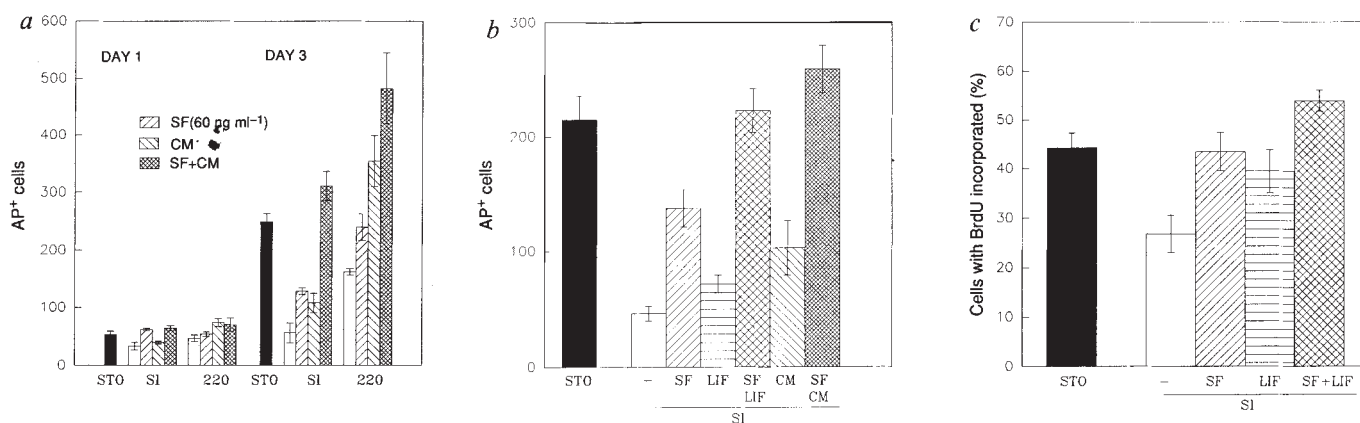


FIG. 3 Effect of Steel factor (SF), LIF/DIA and STO cell conditioned medium on PGC number and proliferation. *a* and *b*, PGCs from 8.5-day-p.c. embryos were cultured with STO cells, homozygous SI/SI⁴ cells lacking the *Sl* gene (indicated by SI), or SI⁴-m220 cells expressing only membrane-associated SF (indicated by 220). STO conditioned medium (CM 100 μ l), 60 ng ml⁻¹ rSF or 20 ng ml⁻¹ LIF were added at the start of the experiment and PGCs counted after 1 and 3 days. (Open bars, no additions). Each experiment was carried out with duplicate wells; numbers are the mean $\pm$ s.e.m. of 3 experiments. Control experiments (not shown) established that 60 ng ml⁻¹ SF and 20 ng ml⁻¹ LIF were in excess under these conditions. *c*, The proportion of PGCs in S phase after 2 days was determined by scoring the number of alkaline phosphatase-positive cells that incorporated BrdU during a 1 h labelling period. Numbers are the mean $\pm$ s.e.m. from 5 experiments.

METHODS. The endothelial-like line, SI/SI⁴, was derived from a long-term

marrow culture established from a 14.5-day-p.c. homozygous *Sl* null embryo liver¹¹. Cells were co-transfected by lipofection with the p γ 3 plasmid encoding hygromycin B phosphotransferase and mouse SF cDNA cloned by reverse-transcribed polymerase chain reaction into the expression vector V19.8 (ref. 8), and selected for hygromycin resistance. The line SI⁴-m220 expresses only the membrane-bound form of SF which lacks the proteolytic cleavage site encoded by exon 6. The cells were used after one and two passages, having been shown by fluorescence-activated cell sorting to express cell-surface-associated SF (50% of cells were positive). STO-conditioned medium was prepared by incubating confluent cultures of irradiated STO cells for 48 h with DMEM plus 10% calf serum. Soluble recombinant human LIF was prepared in *E. coli*. BrdU labelling was carried out using the Amersham cell proliferation kit (RPN 20) and fixing cells in acetic acid:ethanol (1:19).

A consistent observation by us and by others^{4,5} is that PGCs have a finite proliferative capacity *in vitro*, which correlates with the timing of their entry into mitotic arrest or meiosis *in vivo*. One possibility²⁰ is that the expression of the *c-kit*-encoded receptor is downregulated. Alternatively, or in addition, there may be programmed uncoupling of the receptor and the intracellular signalling mechanism²⁶. These, and other alternatives, are being investigated.

Note added in proof: After submission of this manuscript two papers have been published showing an effect of Steel factor on PGC survival *in vitro*^{27,28}.

Voltage-sensing residues in the S4 region of a mammalian K⁺ channel

Emily R. Liman & Peter Hess*

Department of Cellular and Molecular Physiology and Program in Neuroscience, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

Frances Weaver & Gideon Koren

Division of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

THE ability of ion-channel proteins to respond to a change of the transmembrane voltage is one of the basic mechanisms underlying electrical excitability of nerve and muscle membranes. The voltage sensor has been postulated to be the fourth putative transmembrane segment¹ (S4) of voltage-activated Na⁺, Ca²⁺ and K⁺ channels¹⁻⁵. Mutations of positively charged residues within S4 alter gating of Na⁴ and Shaker-type K⁺ channels⁵, but quantitative correlations between the charge or a residue in S4 and the gating valence of the channel have not yet been established. Here, with improved resolution of the voltage dependence of steady-state activation, we present estimates of the equivalent gating valence with sufficient precision to allow quantitative examination of the contribution of individual charged residues to the gating valence of a mammalian non-inactivating K⁺ channel. We conclude that at least part of the gating charge associated with channel activation is indeed contributed by charged residues within the S4 segment.

To test whether a residue is part of the voltage sensor of an ion channel, we must examine whether changes in the charge

Received 3 July; accepted 10 September 1991.

- Ginsburg, M., Snow, M. H. L. & McLaren, A. *Development* **110**, 521-528 (1990).
- Tam, P. P. L. & Snow, M. H. L. *J. Embryol. exp. Morph.* **64**, 133-147 (1981).
- Mintz, B. & Russell, E. S. *J. exp. Zool.* **134**, 207-230 (1957).
- Godin, I., Wylie, C. & Heasman, J. *Development* **108**, 357-363 (1990).
- Donovan, P. J. *et al. Cell* **44**, 831-838 (1986).
- Chabot, B. *et al. Nature* **335**, 88-89 (1988).
- Geissler, E. N., Ryan, M. A. & Housman, D. E. *Cell* **55**, 185-192 (1988).
- Martin, F. H. *et al. Cell* **63**, 203-211 (1990).
- Williams, D. E. *et al. Cell* **63**, 167-174 (1990).
- Zsebo, K. M. *et al. Cell* **63**, 195-201 (1990).
- Zsebo, K. M. *et al. Cell* **63**, 213-224 (1990).
- Huang, E. *et al. Cell* **63**, 225-233 (1990).
- Copeland, N. G. *et al. Cell* **63**, 175-183 (1990).
- Flanagan, J. G. & Leder, P. *Cell* **63**, 185-194 (1990).
- Anderson, D. M. *et al. Cell* **63**, 235-243 (1990).
- Nocka, K. *et al. EMBO J.* **9**, 3287-3294 (1990).
- Smith, A. G. *et al. Nature* **336**, 688-690 (1988).
- Williams, L. *et al. Nature* **336**, 684-687 (1988).
- Orr-Urtreger, A. *et al. Development* **109**, 911-923 (1990).
- Manova, K. & Bachvarova, R. F. *Dev. Biol.* **146**, 312-324 (1991).
- Mitsui, Y., Zsebo, K. & Hogan, B. L. M. *Nature* **347**, 667-669 (1990).
- Schmitt, R. M., Bruyns, E. & Snodgrass, R. H. *Genes Dev.* **5**, 728-740 (1991).
- Nishikawa, S. *et al. EMBO J.* **10**, 2111-2118 (1991).
- Ogawa, M. *et al. J. exp. Med.* **174**, 63-71 (1991).
- Flanagan, J. G., Chan, D. C. & Leder, P. *Cell* **64**, 1025-1035 (1991).
- Rottapel, R. *et al. Molec. cell. Biol.* **11**, 3043-3051 (1991).
- Godin, I. *et al. Nature* **352**, 807-809 (1991).
- Dolci, S. *et al. Nature* **352**, 809-811 (1991).

ACKNOWLEDGEMENTS. We thank P. Donovan of NCI at Frederick and M. Ogawa of Kumamoto University for advice, S. Suggs for constructing the Steel factor expression vector and B. Samal of Amgen for recombinant LIF. This work was supported in part by the NIH (B.L.M.H.).

* To whom correspondence should be addressed.

of that residue correlate with changes of the total gating valence (z_t) of the channel. Each residue is expected to contribute to z_t by an amount equal to its charge multiplied by the fraction of the electric field ('electrical distance', d) the residue traverses during channel activation. z_t can be estimated from the voltage dependence of steady-state activation⁶. Theoretically, at infinitesimal values of activation, the slope of activation curves plotted on a logarithmic ordinate becomes constant (limiting slope) and independent of the number of states, allowing estimation of the total gating valence z_t in a model-independent way⁷. Figure 1 illustrates this method of extracting values of z_t from small fractional steady-state activation. We used a delayed rectifier K^+ channel (RCK1; refs 8–10) because this channel reaches steady-state activation without significant inactivation. Figure 1d shows that the total gating valence z_t obtained from the limiting slope of semi-logarithmic tail activation curves gives a value of $z_t \approx 7$ unitary charges for RCK1 channels. Different models for channel activation predict different shapes of activation curves, but, as expected⁷, converge to a similar limiting slope on semi-logarithmic plots. Measurements are typically possible down to 0.1–1% of maximal activation. A good fit to the entire activation curve is obtained with a model which assumes four subunits and a final voltage independent opening and closing step. This model is consistent with the known subunit stoichiometry of K^+ channels¹¹ and with biophysical data on the mechanism of activation of this¹⁰ and other¹² K^+ channels.

Figure 2b, c shows representative activation curves obtained with mutant RCK1 channels in which either the first or second arginine in the S4 region (R1 and R2, respectively; see Fig. 2a) was mutated to neutral or negatively charged residues (one-letter amino-acid code). The effects on the gating valence are summarized in Fig. 2d, which plots z_t as a function of the relative charge in each subunit. The simplest expectation for mutations of a gating charge is that the value of z_t should reflect the charge but not the chemical side-chain of the mutated residue. Substitution of a positive charge with a negative one (for example, Arg to Asp) should decrease z_t by twice the amount observed with neutralization (for example, Arg to Gln). These expectations are nicely met for mutations at R2, where, except for the conservative mutation to Lys (R2K), channels with equal charge have closely similar values of z_t . The decrease in z_t by 1.5 for each change of charge by -1 can be interpreted to mean that R2 contributes 22% of the channel's gating valence. With 4 identical subunits in a functional channel, the valence of each sensor is decreased in the neutral mutants of R2 by $1.5/4 = 0.38$ charges, which means that residue R2 crosses 38% of the transmembrane electric field during channel activation.

In contrast to charge substitutions at R2, mutations at R1 did not satisfy the simple expectations mentioned. Substitutions of R1 by Gln (R1Q) or Asn (R1N) decreased the value of z_t from 6.8 to 2.8 (R1Q) and 3.3 (R1N). The magnitude of this effect is unexpected, as the decrease in z_t by ~ 4 charges can only be explained if the R1 residues in the four subunits traversed the entire electric field upon activation. Furthermore, reversing the charge at R1 (R1E or R1D) decreased z_t by little more than neutralization of R1. Mutations at R1 thus seem to change not only the charge at that residue, but also the mechanism of activation itself. More direct evidence for this is obtained from a double mutant in which both R1 and R2 have been substituted by the neutral residue Gln (R1QR2Q; Fig. 2c). Contrary to the expectation that the decreases of z_t observed with neutralization of each charge alone should be additive for the double mutant, the value of z_t for R1QR2Q was reduced by less than that of the single mutant R1Q.

We did not observe expression of channels with charge substitutions at the four positive residues in the middle of S4 (R3 mutated to Q, I, L or N; R4 to Q, I, P or L; K5 to Q or I; and R6 to Q or D), whereas substitutions at the seventh positive charge, K7, in S4 (K7Q, K7E, K7D) gave wild-type values of z_t (ref. 13). An attempt to shift the Arg at position 2 by one

residue also failed to produce functional channels. In this mutant R1 and R2 were neutralized to Gln, and the amino acid preceding R2 (Ile 294) was replaced by an Arg.

Mutations that decrease z_t should also decrease the voltage dependences of the individual forward (opening) and backward (closing) rate constants in the activation pathway. The voltage dependence of closing transitions can be estimated from tail currents at tail potentials negative enough to make the voltage-dependent forward transitions negligible (see legend to Fig. 3).

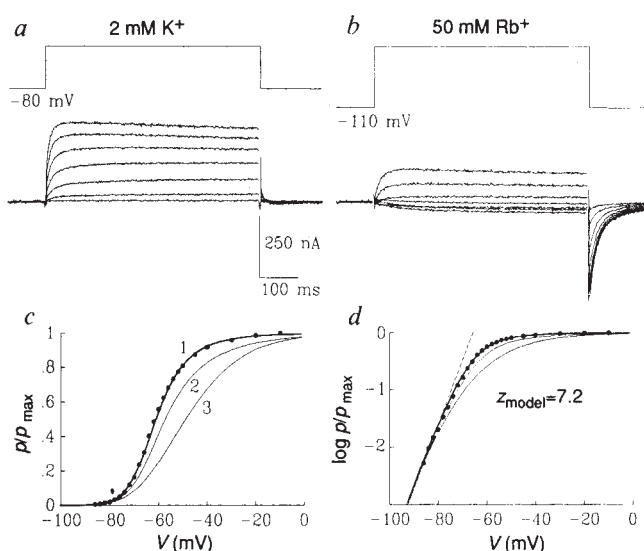
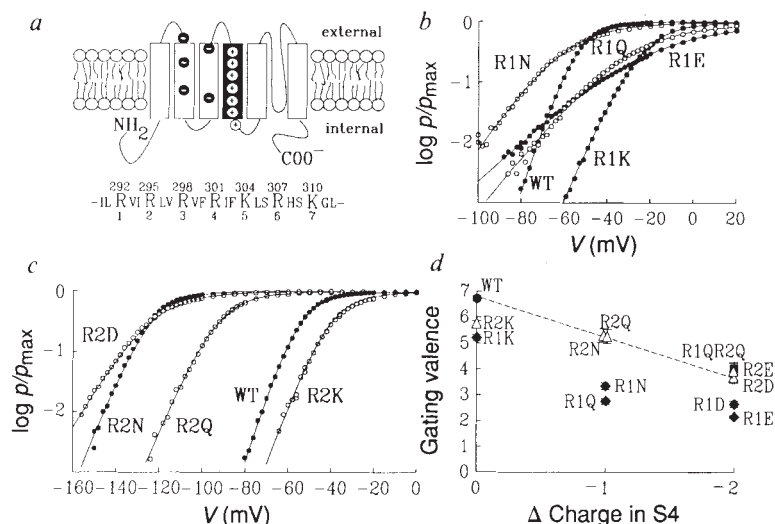


FIG. 1 Estimation of the total gating valence z_t from steady-state activation curves. **a**, K^+ currents in RCK1-channels expressed in a *Xenopus* oocyte and recorded by a two-microelectrode voltage clamp; holding potential -80 mV. Different traces were obtained at test potentials increasing in 10 mV increments between -60 and 0 mV. Linear components of capacity and leak currents were subtracted digitally. Bath solution (in mM): KCl (2), NaCl (118), CaCl_2 (0.3), MgCl_2 (1), HEPES (10), pH 7.2. **b**, Same oocyte as in **a**, after exchange of bath solution with the following (in mM): RbCl (50), NaCl (70), CaCl_2 (0.3), MgCl_2 (1), HEPES (10) pH 7.2. To increase resolution at low values of channel activation, we measured currents through RCK1 channels with 50 mM external Rb^+ as charge carrier to obtain large inward tail currents upon repolarization. Rb^+ slows the rate of channel closure (deactivation)²⁴ and thus further facilitates the measurement of tail currents. Steady-state activation was achieved by 750 -ms-long depolarizing pulses. Holding potential -110 mV, test potentials ranging from -60 to 0 mV in 10 mV increments. **c**, Steady-state activation curve for wild-type RCK1 channel obtained with 50 mM external Rb^+ ('tail activation curve'). Fractional activation ($p/p_{\max}$) was measured as the amplitude of tail currents 3 ms after repolarization to a constant tail potential (isochronal tail measurements), normalized by the maximum tail amplitudes obtained at positive test potentials and plotted as a filled symbol for each test potential preceding repolarization. Smooth curves represent calculated steady-state activation curves for different models in which the total gating valence is kept constant at $z_t = 7.2$. All activation schemes are linear, but differ in the number of closed states preceding a single open state. (1) '4 subunits', 5 closed states, final voltage-independent step with opening rate constant of $30,000$ per s and closing rate constant of $2,200$ per s (ref. 10); (2) '6 subunits', 7 closed states, same voltage independent-last step as in (1). (3) '4 subunits', 4 closed states, no voltage-independent step. Each model curve was shifted to overlap at the foot of the activation curve. **d**, Same as **c**, but plotted on logarithmic ordinate. METHODS. The preparation of *Xenopus* oocytes, injection of oocytes with *in vitro* transcribed RNA and recording of membrane currents by two-microelectrode voltage clamp have been described¹⁰. Currents were recorded by a Dagan 8500 voltage clamp amplifier, filtered (1 – 10 kHz, -3dB), digitized and analysed on a PDP11/73 computer. Special care was taken to avoid voltage errors due to current flow through the bath resistance (1 – $2\text{k}\Omega$ ms with our solutions). We found that even with maximal series resistance compensation (~ 50 – 70%), errors were unavoidable at currents exceeding $5\text{ }\mu\text{A}$. Experiments kept for analysis were limited to peak currents of less than $3\text{ }\mu\text{A}$. The absence of significant endogenous currents was routinely checked by recordings from uninjected oocytes. All experiments were at room temperature (22°C).

FIG. 2 The gating valence of mutants at residues R1 and R2 in S4. *a*, Putative transmembrane folding scheme of voltage dependent K^+ channels. The S4 sequence of RCK1 is shown below the folding scheme. Residue numbers are indicated above Arg (R) and Lys (K) residues. For simplicity we refer to the residues by the numbers indicated below the sequence. The complete sequence of RCK1 can be found in refs 8 and 9. *b*, Semi-logarithmic steady-state activation curves for wild type (WT) and mutants at residue R1 (Arg 291). Recording conditions as in Fig. 1. Solid lines are fits by model (1) in Fig. 1c. *c*, The model curves differ only in z_t and an arbitrary voltage offset. Best fits were determined by eye. *d*, Same as *b*, but for mutants at residue R2 (Arg 294). *e*, Plot of the total gating valence z_t for mutations at R1 and R2. Abscissa plots the change of charge introduced by the mutation. Values for z_t were obtained as described in Fig. 1. $\blacklozenge$, Mutations at R1; $\triangle$, mutations at R2; $\blacksquare$, R1QR2Q. Each symbol shows the mean values $\pm$ s.e.m. of z_t of the following number of measurements (mutation in parenthesis): 16 (wild-type), 4 (R1K), 6 (R1Q), 7 (R1N), 7 (R1E), 7 (R1D), 5 (R2K), 4 (R2Q), 8 (R2N), 3 (R2E), 5 (R2D). Broken line drawn by eye.

METHODS. Single amino-acid substitutions were made using the polymerase chain reaction (PCR). Two PCR products were generated with constant oligonucleotide primers at residues 600–621 and 1,646–6,657 (ref. 8), and two variable abutting oligonucleotides to the S4 region, one of which contained the mutation. The two PCR products were blunt-ended, kinased, digested with *Hind*III and *Clal*, ligated and subcloned as a 1.03-kb insert into RCK1 with a *Clal*–*Hind*III deletion, in pGEM-3Z (Promega). Degenerative oligonucleotides were used to generate multiple mutations at the same residue. The presence of the mutation was



verified by double-stranded dideoxy chain termination sequencing of the entire fragment amplified by PCR. After linearization of the plasmid complementary cDNA by *Hind*III, RNA was transcribed *in vitro* with T7 RNA polymerase.

The *e*-fold change of tail deactivation per 25.5 mV (Fig. 3b) corresponds to an equivalent gating valence $z_c = 0.99$. Figure 3c compares the voltage dependence of tail deactivation for wild-type channels and charge mutations at positions R1 and R2. The same trends are obvious for the values of z_t (compare Figs 2d and 3c). Notably, for mutations at residue R2, z_c is decreased by a similar amount for each –1 change of

the charge at that position. In a pattern resembling that observed with the total gating valence, charge neutralizations at R1 have more pronounced effects on z_c compared with those at R2, and the double mutant R1QR2Q has a higher value of z_c than R1Q alone. The similar patterns obtained for z_t and z_c are good evidence that charge substitutions at R1 or R2 do indeed change the gating valence of the voltage sensor.

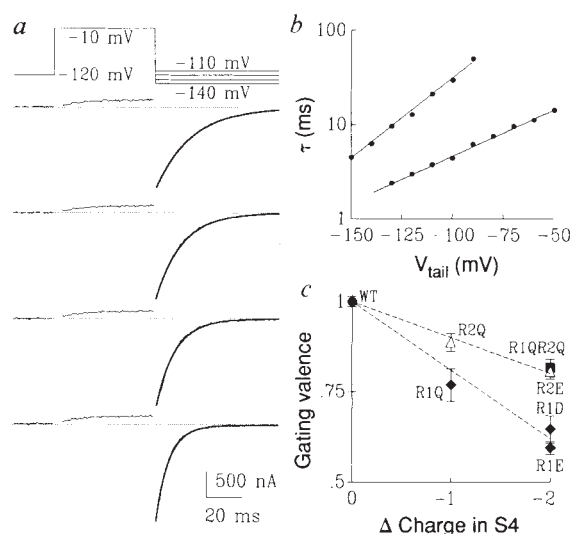


FIG. 3 Voltage dependence of tail deactivation. *a*, Current traces from wild-type channels obtained with the voltage protocol indicated. 50 mM Rb^+ was in the external solution, recording conditions were as in Fig. 1. Time constants of tail deactivation were obtained by maximum likelihood fits of single decaying exponentials to the tail currents (smooth curves superimposed on tail currents in each trace). *b*, Plot of the time constants (τ) of tail deactivation versus tail potential for wild-type channels and mutant R1E. The voltage dependence of tail deactivation was obtained by maximum likelihood fits to an exponential voltage dependence, as indicated by the straight lines in the semi-logarithmic plot. From this fit the voltage change ΔV per *e*-fold change of the tail time constant was used to calculate $z_c = RT/F \times 1/\Delta V$. *c*, Plot of the closing valence z_c for mutations at R1 and R2. Abscissa, change in charge introduced by mutation. $\blacklozenge$, Mutations at R1; $\triangle$, mutations at R2; $\blacksquare$, R1QR2Q. In theory, if channel closing involves a voltage-independent initial step with rate constants k_o (opening) and k_c (closing), followed by a voltage-dependent transition k_b to the next closed state (refs 10, 13), tail deactivation should be biexponential. But calculation of the time constants of tail deactivation with rate constants for the voltage-independent step estimated previously (ref. 10; $k_o = 30,000$ per s; $k_c = 2,200$ per s) shows that over the voltage range of our tail measurements the second exponential component would always be too fast for us to resolve ($< 27 \mu s$), and its amplitude small ($< 6\%$ of the slow component). The time constant of the dominant (measurable) exponential component (τ_m) contains an expression in k_o , k_c and k_b , but, at potentials at which $\tau_m \gg 1/k_c$, τ_m varies exponentially with voltage and the *e*-fold change in τ_m per ΔV can be used directly to obtain the desired voltage dependence of k_b . Thus, as long as our measured time constants vary exponentially with voltage, the presence of voltage independent transitions in the closing pathway should not compromise the measurements of the closing valence. The voltage dependence of closing transition could be obtained more directly from single channel recordings, but this approach is difficult, because most of the voltage dependence is believed to reside in transitions between closed states, and because of the small anticipated effects ($\sim 20\%$ for neutral mutants at R2, see Fig. 2).

All mutations at R1 and R2 produced shifts of the activation curves along the voltage axis (Fig. 4). Shifts of activation curves could result from a number of mechanisms, outlined in the legend to Fig. 4. But for two mutants with equal values of z_i and charge q , and in the absence of evidence for changes in voltage-independent transitions, a difference in the midpoint of activation ($V_{1/2}$) should result solely from changes in the chemical free energies governing the equilibrium between permissive and non-permissive states. Thus, the observation of a 25 mV difference in $V_{1/2}$ between R2Q and R2N (see also the pairs R1E/R1D and R2E/R2D) emphasizes the importance of the precise chemical interactions between S4 residues and the surrounding protein. This finding is consistent with the known ability of mutations even in uncharged residues to cause shifts in $V_{1/2}$ (refs 14–16).

Although unequivocal interpretation of the observed shifts is not possible for any one mutation (see legend to Fig. 4), the pattern of shifts obtained with mutations at R1 and R2 (Fig. 4a) can be interpreted as shown in Fig. 4b. In this scheme, residue R1 interacts favourably with its protein surrounding (here shown for simplicity as a negative countercharge) in the permissive (open) conformation, whereas residue R2 does so mainly in the closed conformation. Thus, neutralization of R1 should destabilize the open state and shift the activation curve positively, whereas the opposite is expected for neutralization at position R2, consistent with our data (Fig. 4a). The same reasoning can be used for positions R3 and R4, if we include the data of Papazian *et al.*⁵, who showed that in *Shaker* K⁺ channels R3Q shifted $V_{1/2}$ positively and R4Q shifted it negatively. According to the model in Fig. 4b, each positive charge in the voltage sensor of S4 interacts favourably with only one countercharge at either the permissive or non-permissive position of the voltage sensor, and at each of the two positions

favourable and unfavourable interactions alternate from one positive charge to the next. This model differs from the one proposed previously (see for example ref. 17), in which each positive charge interacts with a negative charge at each position of the sensor.

Some previous predictions of the tertiary structure of voltage-dependent ion channels have placed S4 charges in close proximity to the channel's permeation pathway^{18,19}. We found no evidence that mutations altered permeation properties of the channel. The reversal potential in 50 mM external Rb⁺, a sensitive measure of K⁺-channel selectivity^{6,20}, was -36.7 ± 2.7 mV (mean $\pm$ s.e.m.; $n=9$) for wild-type channels, and was not statistically different ($P>0.05$) for any of the mutant channels described here (range: -29.1 mV to -41.8 mV, 3 to 9 measurements for each mutant). Single channel conductances, measured with physiological salt concentrations (5.4 mM K⁺ outside, 145 mM K⁺ inside), had similar values for wild-type channels (15.4 ± 0.6 pS, mean $\pm$ s.e.m.) and mutants R2D (13.0 ± 0.5 pS) and R1E (16.1 ± 1.0 pS).

Previous attempts to correlate quantitatively the change of a residue's charge with the effect on the channel's gating valence have been hampered by poor experimental precision⁵, rapid inactivation^{4,5}, which precludes steady-state measurements of activation, and the need for model-dependent interpretations⁴. With improved resolution of activation curves and measurements of the voltage dependence of tail deactivation, we have shown that S4 charges are involved in the voltage-dependent conformational changes occurring during activation of the RCK1 K⁺ channel. Our estimate of the wild-type gating valence is comparable to results from native delayed rectifier and *Shaker* K⁺ channels^{10,12,21,22} and the deduced valence per voltage sensor (1.7 charges) is close to the elementary gating current fluctuations (2.3 charges) presumed to result from movements of a

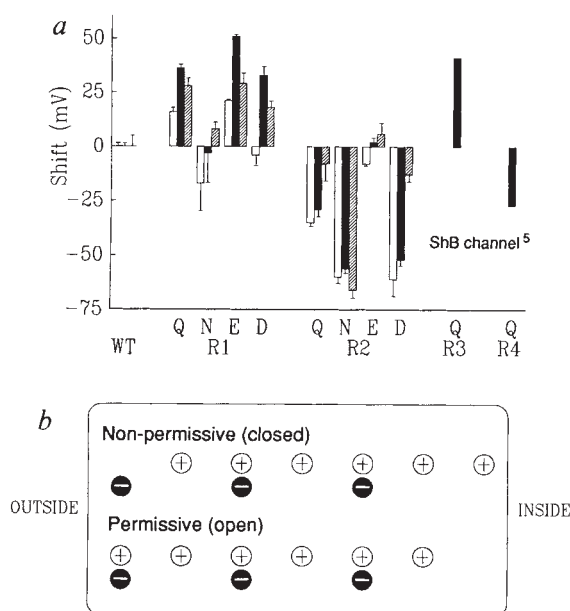


FIG. 4 Shifts of voltage-dependent parameters in mutants at R1 and R2. *a*, For each mutation at R1 and R2, the position of three measured parameters relative to wild-type channels is shown. Open and filled columns: voltage of 10% and 50% maximal steady-state activation, respectively; hatched columns: voltage at which the time constant of tail deactivation is 10 ms. Values are means $\pm$ s.e.m. The double mutant R1QR2Q produced the following voltage shift (relative to wild-type): -33 mV and -23 mV for 10% and 50% maximal activation, respectively. The voltages of 50% activation for R3Q and R4Q are shown for *Shaker* B channels as published in ref. 5. The amino-acid sequences of S4 in *Shaker* B (Sh B) and RCK1 channels are identical. It should be noted that voltage shifts can generally be specified only when the voltage dependence of the two curves is unchanged. Otherwise, shifts vary with the voltage used for comparison. *b*, Schematic representation of postulated interactions between positive and negative charges at the two positions of the voltage sensor. Upon depolarization, the six positive residues move to the left by one position, whereas the negative countercharges are not mobile with respect to the electric field. For simplicity, net negative charge is depicted, but partial negative charge or a negative dipole potential would suffice, as long as an energetically stable interaction with the positive gating charge is achieved. Interpretation of shifts: For a simple two-state equilibrium between a 'permissive' and a 'non-permissive' conformation connected by voltage-dependent forward (k_f) and backward (k_b) rate constants $k_f = k_{f0} \times \exp(z_f V_s F/RT)$ and $k_b = k_{b0} \times \exp(-z_b V_s F/RT)$, the voltage at which half the sensors are in either state corresponds to the midpoint of an activation curve and is given by $V_{1/2} = \ln(k_{b0}/k_{f0}) \times RT/F(z_f + z_b)$. Here, V_s is the voltage difference across the sensor, and k_{f0} and k_{b0} are the values of the rate constants at $V_s = 0$. Surface potentials and intramolecular dipole potentials must influence V_s and thus $V_s \neq V_m$ (V_m is the measured membrane potential). Charge mutations could change the intramembrane potential profile and thus V_s . Unless $k_{f0} = k_{b0}$, changes in z should shift $V_{1/2}$, with the direction and magnitude of the shift depending on the value of V_m at which $V_s = 0$. If voltage-independent as well as voltage-dependent transitions connect closed to open states, the values of the voltage-independent rate constants will also change $V_{1/2}$. The data in *a* are consistent with the model in *b*, but insufficient to prove it. The insignificant shift observed with R2E does not invalidate the model as side-chain-specific chemical interactions, indicated by the large difference between R2E and R2D, could obscure the shifts produced by the change in charge alone.

single gate in Na^+ channels²³. Phenotypes from point mutations may be affected by unforeseen changes in higher-order protein structure: our results for point mutations at R1 indicate that these mutations may do more than simply change the charge on that residue. The unexpectedly large, and somewhat side-chain-specific decreases of z_i and z_c for mutants R1Q and R1N could be interpreted to mean that R1, in addition to carrying part of the gating charge, is important for stabilizing the higher-order structure of the voltage sensor, and that in the absence of R1 the remaining sensor charges are arranged in a different conformation which results in the greatly reduced value of z_i . Interpretation of the mutants at R2 is more straightforward. Except for the mutation to Lys (R2K), there is a simple correla-

tion of z_i and z_c with the charge on the mutated side-chains. This correlation argues that mutant phenotypes at R2 can be attributed mainly to changes of the charge at that residue, and suggests that R2 moves through 38% of the electrical field during channel activation. This conclusion is inconsistent with a model¹⁹ in which R2 (designated at +1 in ref. 19) does not move relative to the electric field during channel activation. We cannot speculate on the precise movement of the voltage sensor. Estimates of the electrical distances covered by other residues would be necessary to evaluate whether all gating charges move by similar amounts, as proposed in the sliding-helix model¹⁷, or whether the gating movement is unequally distributed among gating charges (for example in a propagating helix¹⁹). □

Received 3 May; accepted 5 September 1991.

1. Noda, M. *et al. Nature* **312**, 121–127 (1984).
2. Tanabe, T. *et al. Nature* **328**, 313–318 (1987).
3. Papazian, D., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 749–753 (1987).
4. Stuehmer, W. *et al. Nature* **339**, 597–603 (1989).
5. Papazian, D., Timpe, L., Jan, Y. N. & Jan, L. Y. *Nature* **349**, 305–310 (1991).
6. Hille, B. *Ionic Channels of Excitable Membranes* 1–426 (Sinauer, Sunderland, 1984).
7. Almers, W. *Rev. Physiol. biochem. Pharmac.* **82**, 96–190 (1978).
8. Baumann, A., Grupe, A., Achermann, A. & Pongs, O. *EMBO J.* **7**, 2457–2463 (1988).
9. Christie, M. J., Adelman, J. P., Douglass, J. & North, R. A. *Science* **244**, 221–224 (1989).
10. Koren, G., Liman, E., Logothetis, D., Nadal-Ginard, B. & Hess, P. *Neuron* **4**, 39–51 (1990).
11. MacKinnon, R. *Nature* **350**, 232–235 (1991).
12. Zagotta, W. N. & Aldrich, R. W. *J. gen. Physiol.* **95**, 29–60 (1989).
13. Logothetis, D. E. *et al. Biophys. J.* **59**, 196a (1991).

14. Auld, V. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 323–327 (1990).
15. McCormack, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 2931–2935 (1991).
16. Lopez, G. A., Jan, Y. N. & Jan, L. Y. *Neuron* **7**, 327–336 (1991).
17. Catterall, W. A. *Science* **242**, 50–61 (1988).
18. Benndorf, K. *Eur. Biophys. J.* **17**, 257–271 (1989).
19. Guy, H. R. & Conti, F. *Trends Neurosci.* **13**, 201–206 (1990).
20. Yool, A. J. & Schwarz, T. L. *Nature* **349**, 700–704 (1991).
21. Bezanilla, F., White, M. M. & Taylor, R. E. *Nature* **296**, 657–658 (1982).
22. Gilly, W. M. F. & Armstrong, C. M. *J. gen. Physiol.* **70**, 965–996 (1982).
23. Conti, F. & Stuehmer, W. *Eur. Biophys. J.* **17**, 53–59 (1989).
24. Swenson, R. P. & Armstrong, C. M. *Nature* **291**, 427–429 (1981).

ACKNOWLEDGEMENTS. This work was supported by NIH, the American Heart Association, the Markey Charitable Trust and a Howard Hughes Fellowship (to E.L.). P.H. is an Established Investigator of the American Heart Association.

A cyclic GMP-activated channel in dissociated cells of the chick pineal gland

Stuart E. Dryer & Dori Henderson

Program in Psychobiology and Neuroscience,
Department of Biological Science B-157, Florida State University,
Tallahassee, Florida, USA

PHOTOTRANSDUCTION in the vertebrate retina is dependent in part on a cyclic GMP-activated ionic channel in the plasma membrane of rods and cones^{1,2}. But other vertebrate cells are also photosensitive. Cells of the chick pineal gland have a photosensitive circadian rhythm in melatonin secretion that persists in dissociated cell culture^{3,4}. Exposure to light causes inhibition of melatonin secretion, and entrainment of the intrinsic circadian oscillator^{5,6}. Chick pinealocytes express several 'retinal' proteins, including arrestin⁷, transducin⁸ and a protein similar to the visual pigment rhodopsin⁹. Pinealocytes of lower vertebrates display hyperpolarizing responses to brief pulses of light^{10,11}. Thus it is possible that some of the mechanisms of phototransduction are similar in retinal and pineal photoreceptors. We report here the first recordings of cyclic GMP-activated channels in an extraretinal photoreceptor. Application of GMP, but not cyclic AMP, to excised inside-out patches caused activation of a 15–25 pS cationic channel. These channels may be essential for phototransduction in the chick pineal gland.

Patch-clamp recordings were made from acutely dissociated pineal cells obtained from chick embryos and newly hatched chicks (18-day-old embryo (E-18) to 2-day-old chick (D-2)). Inside-out patches were obtained by withdrawal of the recording micropipette after formation of a stable cell-attached seal. In 15 out of 29 inside-out patches tested, bath application of salines containing micromolar concentrations of cyclic GMP caused activation of ionic channels (Fig. 1a). In this example,

application of 500 μM cGMP activated several identical ionic channels. These channels remained active for as long as cGMP was present (up to several minutes). Bath application of 500 μM cAMP failed to activate channels, but a subsequent application of cGMP again resulted in channel activation.

Cyclic GMP-activated channels reversed close to 0 mV in symmetrical 145 mM NaCl (Fig. 1b). Active channels displayed frequent brief transitions to the closed state, even at saturating concentrations of cGMP. For this reason, the maximum probability of finding channels in the open state (P_0) was always <0.6 . Similar results have been obtained in retinal photoreceptors^{12,13}. Brief closures were more frequent at negative membrane potentials, and poorly resolved transitions may have contributed to the slight rectification apparent in Fig. 1b. Unitary currents had an amplitude of -1.1 pA at a membrane potential of -70 mV and 1.7 pA at $+70$ mV. These values correspond to chord conductances of 16 and 24 pS, respectively.

The continuous activation of the channels in the presence of high concentrations of cGMP made it possible to examine their ionic dependence by application of ramp voltage commands (Fig. 1c). In symmetrical 145 mM NaCl containing 500 μM cGMP, and in the absence of divalent cations, application of voltage ramps resulted in current traces that reversed polarity at close to 0 mV. Bath application of saline containing 500 μM cGMP, 125 mM CsCl, and 20 mM NaCl caused a marked reduction in the current observed at positive membrane potentials. There was also a slight positive shift in the zero-current potential. These results indicate that the cGMP-activated channels are selective for cations, and also show that Na^+ is more permeant than Cs^+ . Similar results were obtained in five other patches, and this result has also been reported in retinal photoreceptors¹⁴. Application of bath salines containing 500 μM cGMP, 145 mM NaCl, and 5 mM MgCl_2 also reduced current, especially at positive membrane potentials (not shown).

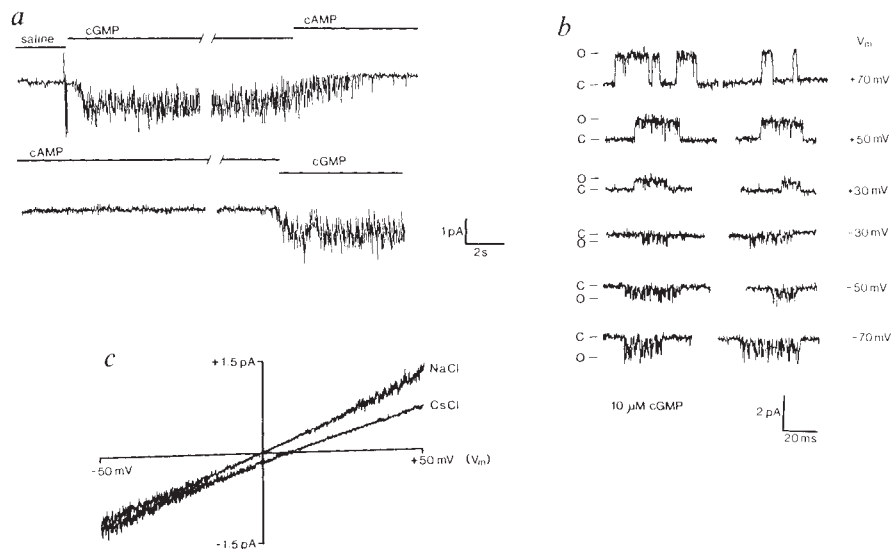
Channel activation was steeply dependent on the concentration of cGMP (Fig. 2). Occasional channel events could be seen with as little as 2 μM cGMP, but were readily detectable in

FIG. 1 Cyclic GMP activates cationic channels in excised inside-out patches. Cyclic nucleotides were applied by total bath perfusion. *a*, Currents in normal saline, and after application of normal saline containing 500 μM cGMP or 500 μM cAMP, as indicated. Note channel activity evoked by cGMP but not by cAMP. Break marks, 1-s gaps in records; membrane potential, -50 mV . In this and in subsequent traces, current flowing into the recording micropipette is shown as outward. *b*, Cyclic GMP-activated channels evoked by 10 μM cGMP at the membrane potentials indicated to the right of the records. Closed (C) and open (O) states are indicated to the left. Note more frequent brief transitions to closed states at negative membrane potentials. *c*, Voltage-dependence of cGMP-activated current revealed by application of ramp voltage commands (-50 to $+50\text{ mV}$ at 100 mV s^{-1}). Continuous channel activity was obtained by bath application of 500 μM cGMP. Data are plotted as the average of five sweeps in normal symmetrical saline (NaCl) consisting of (in mM): 145, NaCl; 10, Na-HEPES; 1, EGTA at pH 7.4, and in CsCl/saline consisting of (in mM): 125, CsCl; 20, NaCl; 10, Na-HEPES; 1, EGTA at a pH of 7.4. Note reduction in cGMP-activated current at positive membrane potentials caused by CsCl/saline, and slight positive shift in the zero-current potential.

METHODS. Pineal glands were dissected from chick embryos or newly hatched chicks (E-18 to D-2). Tissue was incubated at 37°C for 15–20 min in a saline free of divalent cations and containing 1 mg ml^{-1} collagenase (Sigma Type II). Cells were transferred to a serum-free medium, dissociated by trituration, and plated onto glass coverslips coated with concanavalin A.

10 μM cGMP. Half-maximal activation invariably occurred between 10 μM and 50 μM cGMP in six patches tested. Similar results have been reported previously in retinal photoreceptors^{1,2}.

Most patches contained more than one ionic channel. But four patches contained only one cGMP-activated channel, as judged by the nearly continuous unitary current in the presence of 500 μM cGMP. These patches were used for preliminary kinetic analysis (Fig. 3). Cyclic GMP-activated channels opened in bursts separated by many brief closures, suggesting the



Cells were allowed to settle for 1–2 h at 37°C to ensure sufficient adhesion to the substrate. Experiments were done at room temperature (22 – 24°C). Pipette solutions in all experiments were the normal NaCl saline described above. Single-channel currents were filtered at 2–5 kHz with a 4-pole Bessel filter before digitization at 20 kHz, except for the records in *c*, which were digitized at 2 kHz. Voltage commands, data acquisition, and analysis was performed with PCLAMP software (Axon Instruments).

existence of multiple kinetic states. At a membrane potential of -70 mV , and with 10 μM cGMP in the bath, the distribution of uncorrected open times could only be fitted with a double-exponential curve with time constants of 0.92 ± 0.09 and $6.00 \pm 0.48\text{ ms}$ (mean $\pm$ s.e.m.). The distribution of closed times could only be fitted as the sum of two exponential components with time constants of 0.21 ± 0.02 and $1.02 \pm 0.14\text{ ms}$ (mean $\pm$ s.e.m.). The distribution of unitary current amplitudes could be fitted with a single gaussian distribution with a mean of 1.06 pA (s.d. 0.12 pA), corresponding to a unitary chord conductance of

FIG. 2 Concentration-dependence of cGMP-activated channels. Cyclic nucleotides were applied by total bath perfusion. *a*, Inside-out patch containing a single ionic channel was exposed to various concentrations of cGMP, as indicated to the left of the records. Membrane potential was -70 mV . Note increase in channel activity with increasing concentrations of cGMP. *b*, Plot of the probability of finding channels in the open state (P_o) versus the concentration of cGMP. Data points are the means of three patches that contained only one cGMP-activated channel and that were exposed to several concentrations of cGMP. Similar results were obtained in three other patches that contained multiple cGMP-activated channels. Line drawn by cubic spline interpolation. Half-maximal activation always occurred between 10 and 50 μM cGMP.

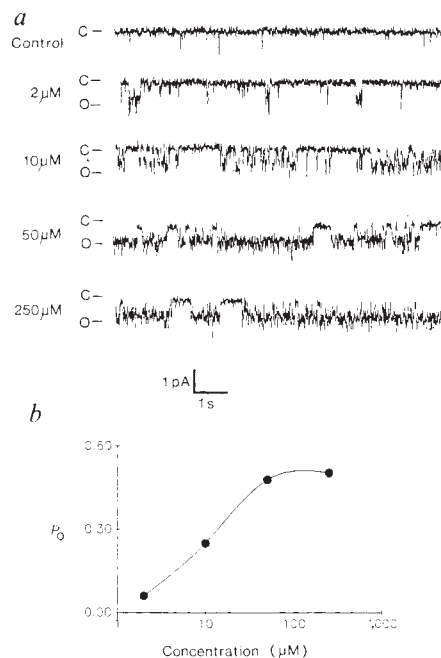
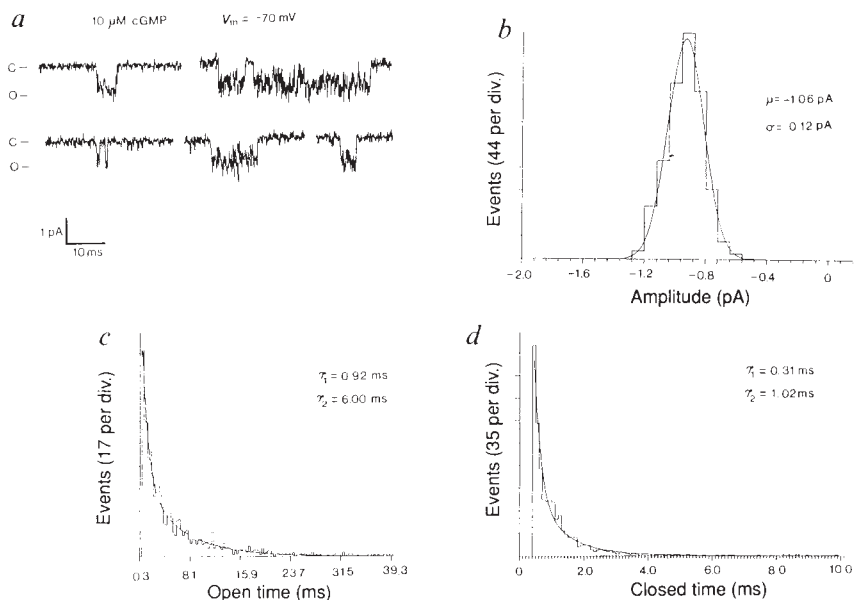


FIG. 3 Kinetic behaviour of cGMP-activated channels. *a*, Examples of raw data used for analysis. This patch contained a single cGMP-activated channel based on the absence of multiple openings in saturating concentrations of cGMP. Data shown was evoked by 10 μ M cGMP. Note frequent brief transitions to the closed state. Membrane potential was -70 mV. *b*, Amplitude histogram of 1,585 events used for analysis. Amplitude histogram is fitted with a single gaussian distribution with mean and standard deviation as indicated. Mean unitary current corresponds to a chord conductance of 15 pS. *c*, Open-time distribution for the same channel. Histogram is fitted with a double-exponential curve with time constants as indicated. *d*, Closed-time distribution for the same channel fitted with a double-exponential curve with time constants as indicated. Note that the presence of occasional long closures (>30 ms) suggests the existence of additional long-lived closed states not apparent from this histogram (see Fig. 2*a* and ref. 15 for examples).

METHODS. Data were filtered at 5 kHz with a 4-pole Bessel filter and stored on videotape for off-line analysis. Data were filtered again at 2.5 kHz with an 8-pole Bessel filter before digitization at 20 kHz. Events were identified using a half-threshold crossing routine that ignored transitions of duration <0.3 ms.



15 pS. Increases in cGMP concentration primarily affected the distribution of closed times (not shown). These characteristics are similar to those of cGMP-activated channels in retinal photoreceptors¹⁵.

We have thus detected a cGMP-activated channel in dissociated cells of the chick pineal gland. These channels are not activated by high concentrations of cAMP, and are therefore different from the cyclic nucleotide-activated channels found in olfactory receptor neurons^{16,17}. Chick pineal cells display two distinct responses to light that seem to be mediated by distinct transduction mechanisms^{5,6}. Light-induced inhibition of melatonin secretion can be blocked by pertussis toxin^{6,18}. This, and the involvement of Ca^{2+} channels in the dark-induced secretion of melatonin¹⁹, suggests that this response could be mediated by cGMP-activated channels. By contrast, light-induced entrainment of the circadian oscillator is resistant to pertussis toxin⁶, and does not seem to depend on changes in membrane potential^{4,18}. This response may use other transduction mechanisms. Finally, it should be noted that chick pineal cells display a circadian rhythm in intracellular cGMP concentration^{18,20}. Therefore it is possible that these channels are also involved in regulating the output of the intrinsic circadian oscillator. □

Plant growth hormones control voltage-dependent activity of anion channels in plasma membrane of guard cells

I. Marten, G. Lohse & R. Hedrich*

Pflanzenphysiologisches Institut, Universität Göttingen, 3400 Göttingen, Germany

THE opening of stomatal pores in the epidermis of plant leaves is caused by an increase in turgor pressure of the guard cells as a result of the accumulation of potassium salts^{1,2}. Although growth hormones have been shown to affect stomatal opening³, the transduction pathways by which growth regulators exert their effects on stomatal action are largely unknown. Here we report that auxins can elicit stomatal opening. These phytohormones modulate anion channels^{4,5} in the plasma membrane in what may be an initial step in regulated volume increase in guard cells. Our patch-clamp experiments demonstrate that auxins can directly interact with the extracellular face of the channel. As a result, its activation potential is shifted towards the resting potential of the cell to favour transient channel opening.

To investigate the primary steps in hormone action we have studied the effect of auxins in the intact epidermis, whole cells and cell-free membrane patches. Anion channels were characterized by their responses to phytohormones and capability to depolarize the membrane potential. When epidermal peels from broad bean (*Vicia faba*) leaves were exposed to external stimuli, stomatal opening was elicited by light, the auxins IAA (indole-3-acetic acid), 1-NAA (1-naphthylacetic acid) and 2,4-D (2,4-dichlorophenoxyacetic acid), as well as the fungal toxin fusaric acid (Fig. 1*a*). Effector-dependent stomatal aperture reached its peak within 5 h after onset of the stimulus. Whereas auxin-treated stomata attained apertures almost similar to the light-

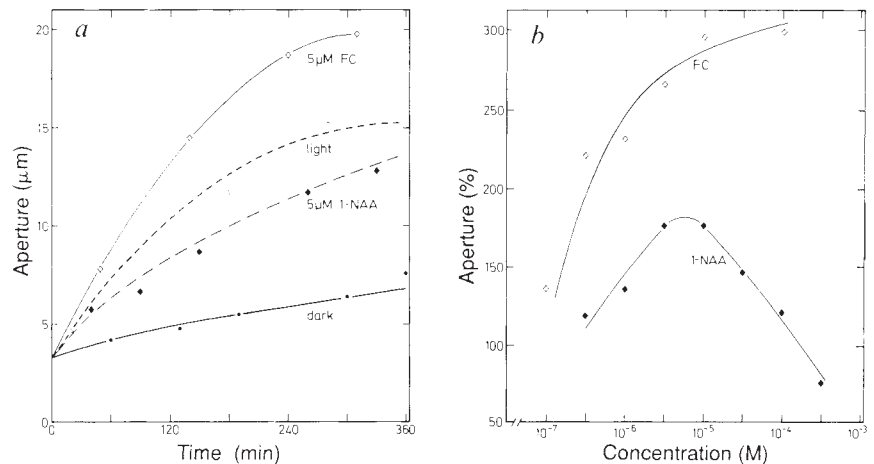
Received 10 June; accepted 27 August 1991.

- Kaup, U. B., Hanke, W., Simmoneit, R. & Lühring, H. *Cold Spring Harb. Symp. quant. Biol.* **43**, 407–415 (1988).
- Yau, K.-W. & Baylor, D. A. *Rev. Neurosci.* **12**, 287–327 (1989).
- Robertson, L. M. & Takahashi, J. S. *J. Neurosci.* **8**, 12–21 (1988).
- Zatz, M., Mullen, D. A. & Moskale, J. P. *Brain Res.* **438**, 199–215 (1988).
- Robertson, L. M. & Takahashi, J. S. *J. Neurosci.* **8**, 21–32 (1988).
- Zatz, M. & Mullen, D. A. *Brain Res.* **453**, 63–71 (1988).
- Collin, J. P. *et al. Neuroscience* **19**, 657–668 (1986).
- van Veen, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 912–917 (1986).
- Deguchi, T. *Nature* **290**, 706–709 (1981).
- Pu, A. & Dowling, J. E. *J. Neurophysiol.* **46**, 1018–1038 (1981).
- Tomotsu, S. & Morita, Y. *J. comp. Physiol.* **159**, 1–5 (1986).
- Matthews, G. & Watanabe, S. I. *J. Physiol., Lond.* **403**, 389–405 (1988).
- Haynes, L. W. & Yau, K.-W. *J. Physiol., Lond.* **429**, 451–481 (1990).
- Menini, A. *J. Physiol., Lond.* **424**, 167–185 (1990).
- Matthews, G. *J. Physiol., Lond.* **403**, 389–405 (1988).
- Nakamura, T. & Gold, G. H. *Nature* **325**, 422–424 (1987).
- Dhallan, R. S., Yau, K.-W., Schrader, K. A. & Reed, R. R. *Nature* **347**, 184–187 (1990).
- Takahashi, J. S. *et al. Hormone Res.* **45**, 279–352 (1989).
- Harrison, N. & Zatz, M. *J. Neurosci.* **9**, 2462–2467 (1989).
- Wainwright, S. D. & Wainwright, L. K. *J. Neurochem.* **413**, 358–371 (1988).

* To whom correspondence should be addressed.

FIG. 1 Stomatal opening in response to 1-NAA, fusicoccin (FC) and light treatment. *a*, Time course of increase in stomatal aperture during stimulation. *b*, Concentration-dependence of auxin and fusicoccin-induced stomatal opening.

METHODS. The lower epidermis of *Vicia faba* leaves, held in the dark for 14–16 h, was peeled off and incubated for up to 5 h in a solution containing 50 mM KCl, 2 mM MES/Tris, pH 6.0. For auxin and fusicoccin experiments, epidermal peels were also incubated under dark conditions. Samples were analysed for stomatal aperture at the times indicated. Each data point in *a* and *b* represents the mean of >100 measurements.



treated ones, fusicoccin exceeded the light control (Fig. 1*a*). Dose-response curves for auxins had an optimum at $\sim 5 \times 10^{-6}$ – 10^{-5} M, with a decay at higher concentrations. Fusicoccin action, on the other hand, is characterized by saturation-type behaviour (Fig. 1*b*). These experiments indicate that guard cells possess recognition sites for auxins as well as mechanisms for translating the hormonal signal into ion fluxes which, in turn, cause volume increase and stomatal aperture. The classical

electrical response of the plasma membrane to auxins that precede growth induction in various plant tissues is characterized by a transient depolarization followed by a sustained hypolarization^{6,7}.

To study the early events in auxin-induced alterations in plasma membrane potential and transmembrane ion fluxes, we applied the patch-clamp technique⁸ to guard cell protoplasts⁹. Patch pipettes were sealed onto guard cell protoplast isolated

FIG. 2 Modulation of the plasma membrane anion channel of guard cells by the auxin 1-NAA showing the time course of auxin-induced inhibition of inward Cl^- current, modulation of peak amplitude of outward Cl^- current and shift in activation potential. *a*, Current-voltage relation of whole-cell anion fluxes in the absence (I), and 15 s and 115 s (II and III) after application, of 10 μM 1-NAA resulting from 1-s voltage ramps from -200 to +60 mV. *b*, Current-voltage relation of anion fluxes in the absence (I), and 15 s and 45 s (II and III) after application, of 100 μM 1-NAA. *c*, Concentration-dependence of 1-NAA-induced shift in activation potential of outward anion channels. Note that potential shift increases logarithmically with 1-NAA concentrations of 5 μM ($n=4$), 10 μM ($n=5$), 25 μM ($n=7$) and 100 μM ($n=13$) (inset). Peak outward chloride fluxes as high as 3 nA were recorded in the range of -40 to -30 mV. Cl^-_{out} currents declined through the decrease in the electrochemical driving force when stepping the membrane potential to depolarized values of more than -30 mV, reversing direction at the Nernst potential for chloride (18 mV). At potentials 40–60 mV positive to E_{Cl^-} , inward chloride currents neither peaked nor saturated but increased, often in a sigmoidal manner (Fig. 2*a*, I; ref. 5). Comparison of whole-cell and single-channel ion fluxes demonstrated that the absence of Cl^-_{out} currents on hyperpolarization is due to a deactivation caused by an increase of the closed times of the anion channel⁵.

METHODS. Guard cell protoplasts were enzymatically isolated from 2–3-week-old leaves of *V. faba*⁵. Patch pipettes were sealed against the plasma membrane to study ion fluxes in the whole-cell configuration and in outside-out patches⁸. Current measurements were made with an EPC-7 patch-clamp amplifier (List-electronic Darmstadt) low-pass filtered with an eight-pole Bessel filter. Data were digitized (VR10, Instrutech), stored on hard disc or video tape and analysed using patch-clamp software (Instrutech) on a Mega Atari ST4. Solutions in Figs 2–5 were composed of 40 mM CaCl_2 , 2 mM MgCl_2 , 10 mM MES/Tris, pH 5.6, in the bathing medium and 150 mM KCl, 2 mM MgCl_2 , 10 mM Mg-ATP and GTP, 0.1 mM EGTA, 10 mM HEPES/Tris, pH 7.2, in the pipette (cytoplasm).

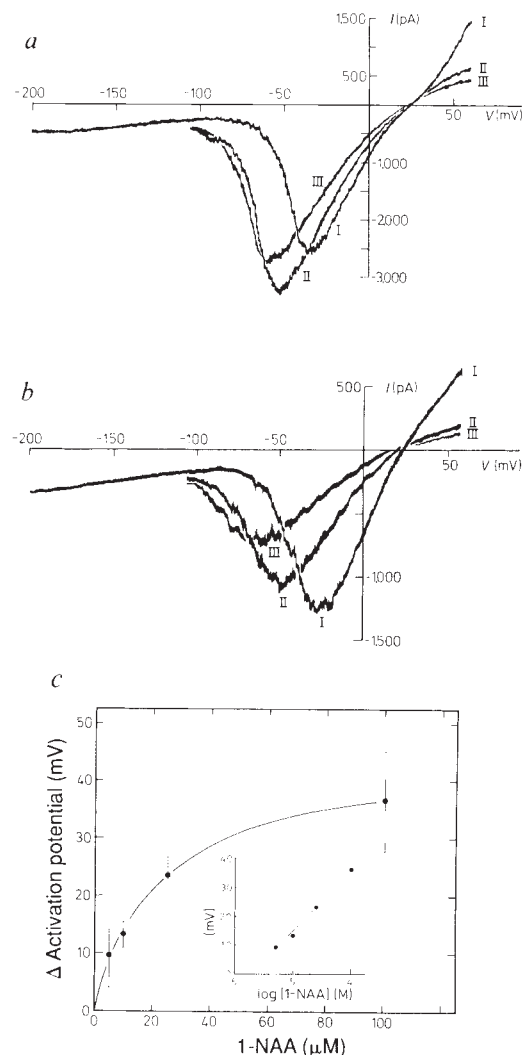
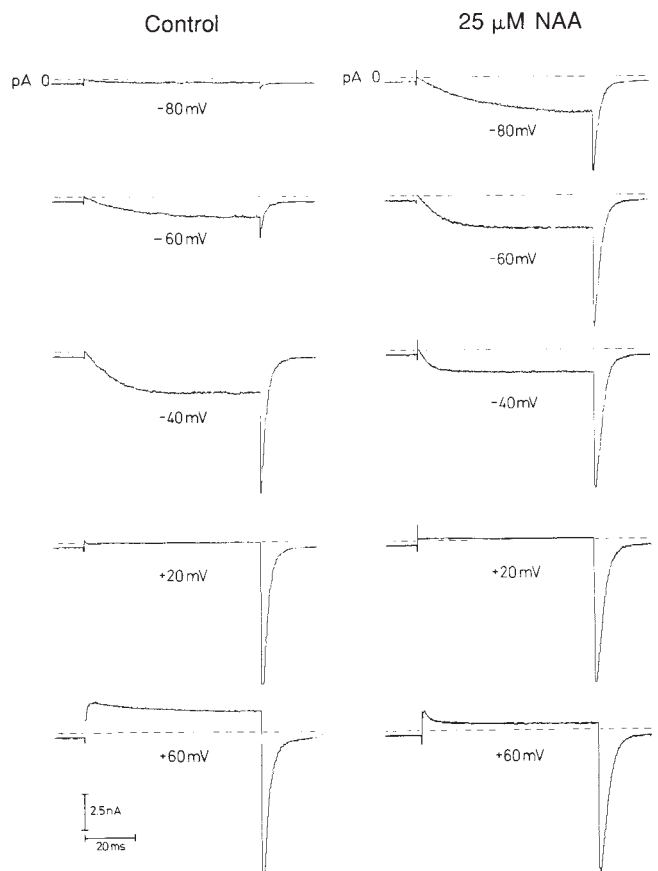


FIG. 3 Modulation of the voltage-dependent kinetics of the anion channel in whole cells. Time course of activation in the absence (left panel) and presence (right panel) of 25 μM 1-NAA. Rise-times of activation were obtained by 70 ms voltage steps to -80 , -60 , -40 , 20 , and 60 mV from a holding potential of -200 mV. Currents were activated with a rise time of about 10.1 and 19.1 ms (at -40 and -60 mV) in the absence of 25 μM NAA and of about 4.9 and 9.5 ms in the presence of 25 μM NAA. These rise times reflect the activation of the anion current at the peak current potential for the two experimental conditions.



enzymatically from auxin-responsive epidermal layers⁵. Patch-clamp recordings were performed in the whole-cell configuration and in outside-out membrane patches⁸. An increase in pump current through the plasma membrane (H^+)ATPase at 3–10 min after auxin application (not shown), which is well documented in other cell types^{6,7}, indicated that hormone sensitivity is maintained during protoplast isolation.

We then concentrated on the direct modulation of plasma membrane anion channels right after extracellular application of the phytohormone. In the whole-cell configuration, voltage-dependent anion channels⁴ are activated by nucleotides and calcium⁵; under these conditions (see figure legends) outward and inward chloride currents indicate that anion channels are the predominant ionic conductances (Fig. 2*a* and *b*). When auxins are applied to the bathing solution the activation potential and current amplitudes are altered in a dose-dependent manner (Fig. 2*a–c*, for voltage protocols see legend). Inhibition of Cl_{in}^- current is observed before peak Cl_{out}^- current and the activation potential shifts to more hyperpolarized potentials (Figs 2 and 3). Kinetic analysis of the whole-cell currents deduced from single voltage-pulses (Fig. 3) reveals similar activation times at the peak current potential in the absence ($t_{1/e}^{-40\text{mV}} = 10.1$ ms) and presence of auxins ($t_{1/e}^{-60\text{mV}} = 9.5$ ms). The shift in voltage-dependence and inhibition of Cl_{in}^- increases with the auxin concentration and incubation time (Fig. 2*a–c*). The shift in activation potential increases logarithmically with 5–100 μM 1-NAA (Fig. 2*c* inset). The amplitude of the shifted outward chloride currents remains unchanged or increased when exposed to 5–25 μM auxin (Fig. 2*a*). But exposure to auxin concentrations above 25 μM inhibits peak outward chloride flow (Fig. 2*b*). The modulation of the anion channel is fully reversible right after removal of auxins from the extracellular solution.

Extracellular auxins are transported into the cytoplasm in the presence of a pH gradient¹⁰; we therefore tested whether the

anion channels are modulated through auxins at the extracellular face of the channel or by cytoplasmic hormone levels. When the cytoplasm was equilibrated with up to 100 μM auxin through the patch pipette, anion currents were not effected until the auxin in question was applied to the bathing solution, indicating that the recognition site for auxins is located at the external face of the channel. Our observation that auxins can modulate the anion channel in the absence of a pH gradient (results not shown) across the plasma membrane is in good agreement with this finding.

To determine whether auxin modulation of the anion channel is direct or if cytoplasmic factors are involved, cell-free membrane patches (outside-out) were exposed to auxin-containing medium. Figure 4*a* shows fluctuations in open-closed transitions of up to three 32-pS anion channels with and without IAA (right panel). Single-channel analysis (Fig. 4*b*, bottom) indicates that auxins shift the activation curve of the channel by -21 mV (± 5 mV; $n = 6$) more negative and transform the long open intervals into a flickering mode reminiscent of a flickering block demonstrated for several channel blockers¹¹. A comparison of the whole-cell currents and single-channel activity ($N \times p$) in a membrane patch excised from the same cell (Fig. 4*b*) indicates that macroscopic currents are generated by auxin-sensitive anion channels. It also shows that cytoplasmic factors are not required for hormone-targeting through the anion channel.

To determine the specificity of the auxin response we tested benzoic acid, fusicoccin and other phytohormones with respect to the modulation of the anion channel. The channel recognized biologically active auxins when applied at 100 μM with a sequence of shift in activation potential of 1-NAA (-36.6 ± 8.7 mV; $n = 13$) > 2,4-*D* (-34.2 ± 5.0 mV; $n = 4$) > IAA (-20.6 ± 5.0 mV; $n = 6$). Fusicoccin, which causes hyperpolarization without a preceding depolarization⁶, is a potent effector of stomatal opening (Fig. 1) but is ineffective in the modulation

of the anion channel. Benzoic acid and the phytohormones abscisic acid, gibberellic acid, or the cytokinin zeatin do not alter the anion channel either.

In cells with resting potentials negative to E_{K^+} , inward-rectifying potassium channels (K_{in}^+) (refs 12, 13) have a depolarizing capacity when activated, similar to anion efflux channels. We therefore investigated the behaviour of K_{in}^+ channels during auxin action on the plasma membrane anion channels. In the presence of external 40 mM $CaCl_2$ and 150 mM KCl in the pipette, inward currents were carried by chloride efflux only, as characterized by the inward hump between E_{Cl^-} (18 mV) and -100 mV, and a high-resistance range negative to -100 mV. On replacement of $CaCl_2$ by KCl, inward K^+ currents were elicited at voltages negative to -100 mV. K^+ currents peaked at around -180 mV (Fig. 5a). At potentials negative to -200 mV, K_{in}^+ currents decay because of a voltage-dependent block by external Ca^{2+} ions¹⁴. Voltage-ramps from -200 or -250 mV to 60 mV from a holding potential of -100 mV allowed us to record the activity of both K_{in}^+ and Cl_{out}^- channels simultaneously. On application of auxins, chloride currents are modulated as expected (Fig. 5; compare Fig. 2), whereas potassium currents are unaltered ($n > 10$). This indicates that K_{in}^+ channels of guard cells do not have target sites for auxins. K^+ channels therefore do not seem to be involved in auxin-induced membrane depolarization.

Our studies show that guard cells respond to auxins by volume

increase (Fig. 1). These plant growth regulators are able to specifically modulate the anion channel and activate the H^+ pump in the plasma membrane, resulting in a change in membrane polarization. Alterations in membrane properties mediated by these two transport proteins are independent of each other and can be separated. When auxins reach the extracellular surface of the plasma membrane the activation potential of Cl_{out}^- is shifted towards the resting potential of the cell (usually -100 to -160 mV). This shift in activation potential and peak potential of the anion channel towards hyperpolarized values, a voltage range where K_{out}^+ channels are not active¹², results in the opening of anion channels causing membrane depolarization rather than salt efflux. Depending on the auxin dose and auxin type, the shift in activation potential is superimposed by an increase or decrease in peak current amplitude. This graduated response allows a high degree of sensitivity to auxin doses by integrating information gained from the magnitude and time course of auxin-induced changes in membrane conductance. When, in the presence of auxin, the resting potential of the cell is within the activation range of the Cl_{out}^- current for a prolonged period, chloride channels inactivate with a half-time of about 10–12 s (ref. 5). The decay of outward chloride current resulting from channel inactivation simultaneously increases membrane resistance. As a result, auxin-stimulated (H^+)ATPases shift the membrane potential into the activation range of potassium influx channels (Fig. 5), allowing long-term accumulation of K^+

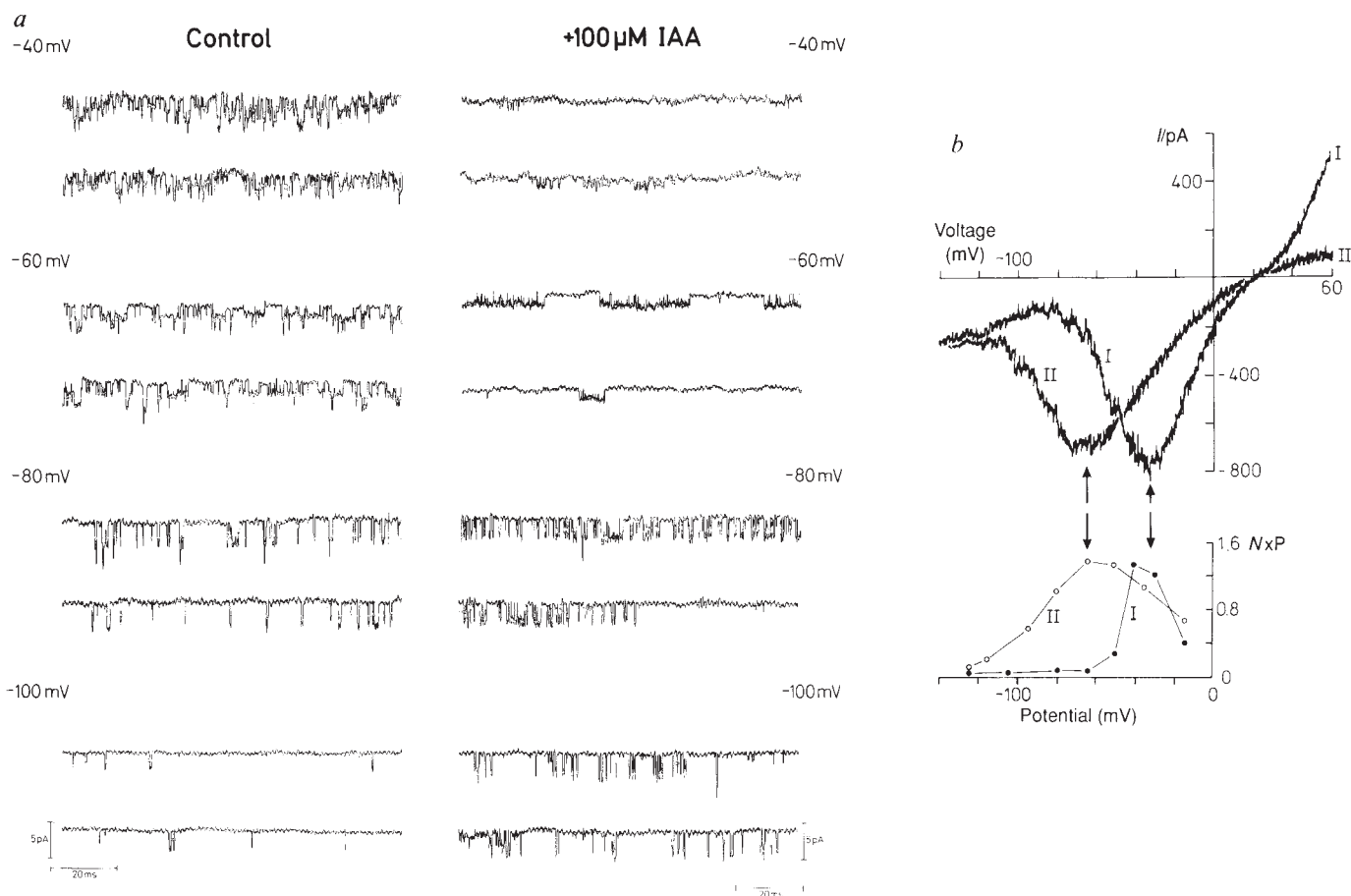


FIG. 4 *a*, Modulation of the properties of single anion channels by auxin. Single-channel activity recorded from an outside-out membrane patch in the absence (left) and presence of 100 μM IAA (right) with the membrane potential clamped to -100 , -80 , -60 , and -40 mV. *b*, Comparison of whole-cell anion currents and single-channel activity before (I) and after

auxin stimulation (II). Single-channel activity ($N \times p$) is expressed by number (N) of channels in the membrane patch and their open probability (P). Note that single auxin-modulated anion channels represent the molecular equivalents for the observed macroscopic anion fluxes.

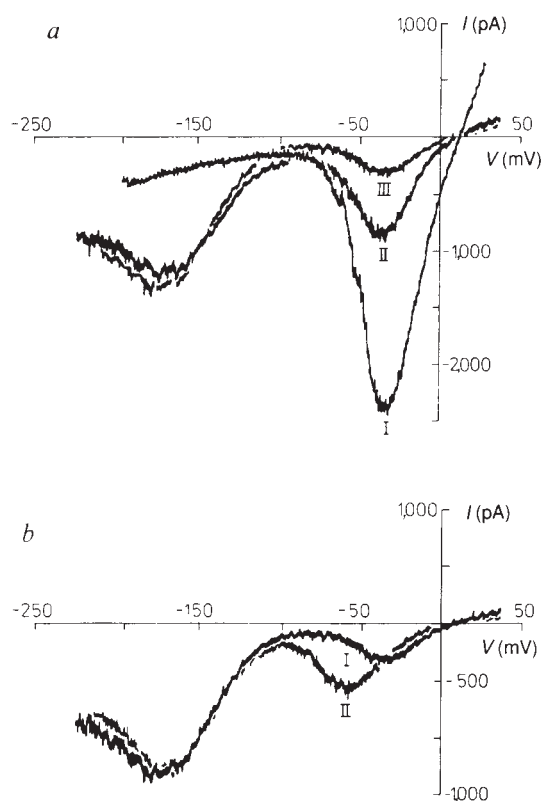


FIG. 5 Insensitivity of inward K^+ channels to auxins. *a*, Replacement of extracellular Ca^{2+} by K^+ causes a decrease in the anion current and an increase of the inward K^+ current. *b*, Simultaneous recording of voltage-dependent potassium and anion channels in the absence (AI and BI) and presence (BII) of 100 μ M IAA. Current-voltage relations of K^+ and anion channels resulted from 1-s voltage ramps from a holding potential of -200 mV (AI) or -120 mV (AII, III and B). Note that potassium currents were not affected by auxin. Bath solutions in *a* were 40 mM $CaCl_2$, 2 mM $MgCl_2$, 10 mM MES/Tris, pH 5.6, and in *b*, 30 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 10 mM MES/Tris, pH 5.6. Pipette solutions were identical to those mentioned in the legend to Fig. 2.

salts^{15,16}. Evidence is accumulating that the (H^+)ATPase is activated by auxins through a complex signalling pathway which is supposed to include several receptor sites and coupling proteins¹⁷, but to our knowledge this is the first evidence for a plant ion channel directly gated by extracellular growth hormones. □

Received 17 July; accepted 28 August 1991.

1. Raschke, K. *Encyclopedia of Plant Physiology* Vol. 7 383–441 (Springer, Berlin, 1979).
2. MacRobbie, E. A. C. *Bot. Acta* **101**, 140–148 (1988).
3. Snaith, P. J. & Mansfield, T. A. *J. exp. Botany* **33**, 360–365 (1982).
4. Keller, B. U., Hedrich, R. & Raschke, K. *Nature* **341**, 450–452 (1989).
5. Hedrich, R., Busch, H. & Raschke, K. *EMBO J.* **9**, 3889–3892 (1990).
6. Bates, G. W. & Goldsmith, M. H. *Planta* **159**, 231–237 (1983).
7. Felle, H., Peters, W. & Palme, K. *Biochim. biophys. Acta* (in the press).
8. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
9. Schroeder, J. I., Hedrich, R. & Fernandez, J. M. *Nature* **312**, 361–362 (1984).
10. Rubery, P. H. & Sheldrake, A. R. *Planta* **118**, 101–121 (1974).
11. Hille, B. *Ionic Channels* (Sinauer, Sunderland, Massachusetts, 1983).
12. Schroeder, J. I., Raschke, K. & Neher, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4108–4112 (1987).
13. Blatt, M. R. *J. Membr. Biol.* (in the press).
14. Busch, H., Hedrich, R. & Raschke, K. *Plant Physiol.* **93**, 17 (1990).
15. Hedrich, R. & Schroeder, J. I. *A. Rev. Plant Physiol.* **40**, 539–569 (1989).
16. Schroeder, J. I. & Hedrich, R. *Trends Biochem. Sci.* **5**, 187–192 (1989).
17. Palme, K. *Int. Rev. Cytol.* **132**, 223–283 (1992).

ACKNOWLEDGEMENTS. This work is dedicated to J. I. Schroeder and M. Spors (9 August 1991). We thank H. Depta (Göttingen) for discussion, R. Hertel (Freiburg), E. Neher (Göttingen) and D. G. Robinson (Göttingen) for comments on the manuscript, and K. Raschke and D. G. Robinson for laboratory facilities. This work was supported by grants from DFG to R.H.

Structure of domain 1 of rat T lymphocyte CD2 antigen

Paul C. Driscoll*, Jason G. Cyster†, Iain D. Campbell*, & Alan F. Williams†

* Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

† MRC Cellular Immunology Research Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK

THE CD2 antigen is largely restricted to cells of the T-lymphocyte lineage and has been established as an important adhesion molecule in interactions between human T lymphocytes and accessory cells¹. In the adhesion reaction, CD2 on T cells binds to LFA-3 on other cells, with binding through domain 1 of CD2 (ref. 2). CD2 can also be a target for the delivery of mitogenic signals to T lymphocytes cultured with combinations of anti-CD2 antibodies^{3,4}. Two predictions that are contradictory have been made for the structure of CD2 domain 1. One suggests an immunoglobulin (Ig) fold, on the basis of sequence patterns conserved in the Ig-superfamily (IgSF)⁵, whilst the other proposes a pattern of alternating α -helices and β -strands, on the basis of

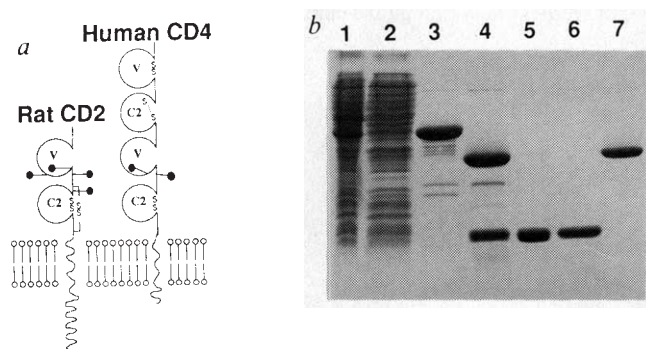


FIG. 1 *a*, Schematic representations for CD2 and CD4 at a cell surface. The circles indicate IgSF domains and the $\uparrow$ symbols, N-linked carbohydrate structures. *b*, Expression and purification of rat CD2 domain 1. Lanes: 1, soluble lysate from *E. coli* expressing the fusion protein of GST-CD2 domain 1; 2, lysate passed through a glutathione agarose affinity column; 3, purified fusion protein; 4, fusion protein cleaved with thrombin; 5 and 6, CD2 domain 1 after purification by affinity chromatography followed by gel filtration; 7, GST.

METHODS. Nucleotide sequence encoding rat CD2 protein residues 1–99 was isolated from the complementary DNA⁷ by polymerase chain reaction (PCR) with the introduction of a *Bam*H1 restriction site immediately preceding the codon for protein residue 1 and a stop codon after the codon for amino-acid 99 followed by an *Eco*R1 site. The PCR product was inserted into the pGEX-2T vector and the sequence was confirmed using oligonucleotides specific for pGEX. The construct was transformed into *E. coli* BL21, a prototrophic strain. For fusion protein production, cells were grown overnight, diluted 1:10 in fresh medium and grown for 1 h at which point IPTG (isopropylthiogalactoside) was added to 0.12 mM, and the cells grown for a further 3 h. The cells were pelleted and resuspended in 1/60 volume 25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.2 mM PMSF, 2 mM iodoacetamide acid and sonicated for 3 $\times$ 30 s using an MSE sonicator on amplitude setting 8. Triton X-100 was added to a final concentration of 1% and the lysate spun at 30,000g for 30 min. The supernatant was passed over a glutathione-agarose column which was then washed with 25 mM Tris-HCl pH 7.4, 140 mM NaCl and eluted with 5 mM glutathione, 50 mM Tris-HCl, pH 8. The eluted material was mixed with 0.05 vol 3 M NaCl, 50 mM $CaCl_2$ and 1 NIH u of bovine thrombin per 2 mg protein. Cleavage was complete after 60 min and the CD2 domain was then purified on an OX34 monoclonal antibody affinity column⁷ followed by gel filtration using a 200 ml Sephadex G75 column. For ¹⁵N-labelling, cells were grown in M9 medium (10 mM $MgSO_4$, 1 mM $CaCl_2$, 0.4% glucose, 48 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 9 mM NaCl) and ¹⁵NH₄Cl (MSD Isotopes, Merck Frost, Canada) at 9.7 mM for ~18 h until the absorbance at 550 nm (A_{550}) was >1.0. The cells were then diluted 1:10 and grown for 2 h until A_{550} ~0.5 when IPTG was added and growth continued for a further 3–4 h.

secondary structure predictions⁶. Thus CD2 domain 1 is an important test case for the validity of IgSF assignments based on sequence patterns. We report here the expression of domain 1 of rat CD2 in an *Escherichia coli* expression system and have determined a low-resolution solution structure by NMR spectroscopy.

A schematic model proposed for the domain structure of CD2 is shown in Fig. 1a together with that for CD4 whose extracellular region may have been derived in evolution from a CD2-like structure by gene duplication of a two-domain segment^{7,8}. Domain 1 of rat CD2 was considered to encompass amino-acid residues 1–99 and this sequence was expressed as a fusion protein with glutathione-S-transferase (GST) of the blood fluke *Schistosoma mansoni*⁹ in *E. coli* with the thrombin cleavage sequence, LVPRGS, between GST and CD2 domain 1. Thrombin cleaves after the Arg(R) residue in this sequence and thus the cleaved product contains the extra residues Gly Ser at the amino terminus of the CD2 domain 1 sequence. The fusion protein was expressed at 40 mg l⁻¹ and was soluble in the lysate of *E. coli* cells, allowing purification on a glutathione agarose column (Fig. 1b, lanes 1–3). Cleavage by thrombin was effective and the CD2 domain was purified by affinity

chromatography followed by gel filtration (Fig. 1b, lanes 4, 5 and 6). The CD2 domain was judged to be folded correctly because it bound to two noncompetitive monoclonal antibodies against rat CD2 (OX34 and OX55) (ref. 4) with an affinity equal to that of the extracellular domain of CD2 expressed in Chinese hamster ovary cells (data not shown). The fusion protein was also expressed in *E. coli* cells grown with ¹⁵NH₄Cl as the sole nitrogen source to yield uniformly ¹⁵N-labelled CD2 domain 1, suitable for multidimensional heteronuclear (NMR) studies.

The CD2 domain 1 preparations yielded excellent NMR spectra and the assignment of the ¹H resonances to individual amino acids was done using a combination of three-dimensional ¹H nuclear Overhauser enhancement ¹⁵N-¹H heteronuclear multiple quantum coherence (NOESY-HMQC) and ¹H homonuclear Hartmann-Hahn-¹⁵N-¹H-HMQC (HOHAHA-HMQC) NMR spectroscopy^{10,11}. Side-chain assignments were obtained using two-dimensional ¹H NMR of an unlabelled sample of the protein in D₂O solution.

Three-dimensional structures for CD2 domain 1 were calculated on the basis of the NOE data, slowly exchanging amide NH groups and ³J_{NH α} spin-spin coupling constants. The

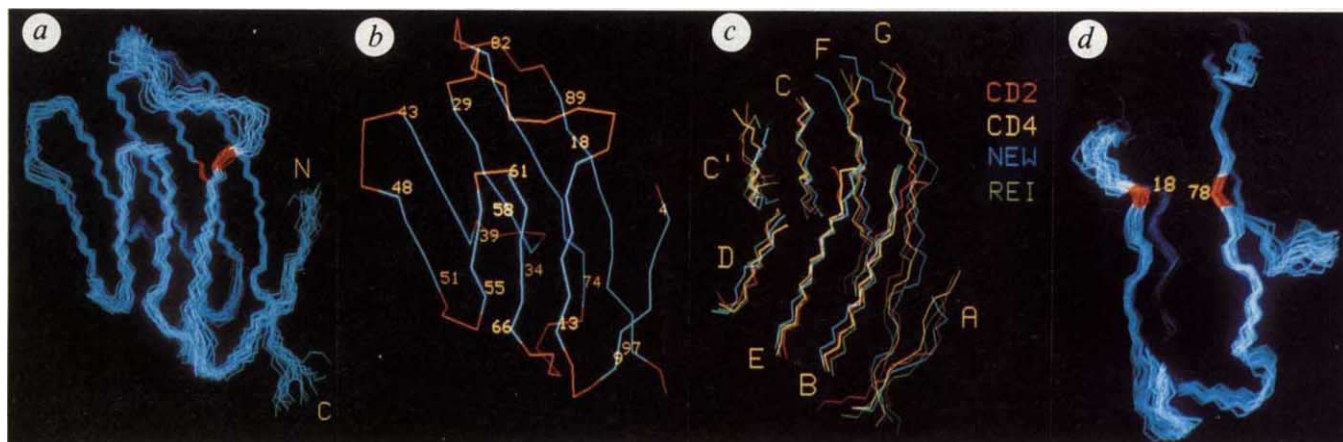


FIG. 2 a, Best-fit superposition of the 16 structures computed for CD2 domain 1 based on 713 NOE distance restraints (20 intraresidue, 346 sequential (residue i , residue j ; $j = i + 1$), 102 short range (i, j ; $j \leq i + 5$), 60 medium range (i, j ; $j \leq i + 10$) and 185 long range (i, j ; $j > i + 10$), 64 ϕ dihedral angle restraints and 68 H-bond distance restraints. Backbone N, C and C α atoms of residues 5–98 are shown. NMR spectra were recorded using a 3 mM pH4.2 CD2 domain 1 sample, at 500 or 600 MHz ¹H frequency and at 23 °C or 30 °C. Distance restraint upper limits were categorized according to the estimated intensity of the NOE cross peak in the spectrum: strong 2.7 Å; medium 3.5 Å; weak 5.0 Å. Appropriate corrections were made to the upper limits in the cases of NOEs connecting degenerate proton resonances. Sequential distance restraints were estimated on a 10-point sliding scale from the integrated cross-peak intensities of the three-dimensional ¹⁵N-¹H NOESY-HMQC NMR experiment (mixing time = 150 ms). In all cases the lower limit of the distance restraint was given by the van der Waals' contact distance. ³J_{NH α} spin-spin coupling constants were measured by lineshape fitting the traces from a two-dimensional ¹⁵N-¹H HMQC- J experiment²⁷. For ³J_{NH α} smaller than 6 Hz the corresponding ϕ torsion angle was restrained in the range -20° to -105°; for ³J_{NH α} larger than 7.5 Hz the ϕ angle was restrained to the range -70° to -120°. H-bonded amide NH groups were identified on the basis of the presence of an NH resonance 2 h after dissolving the protein in D₂O solution. The distance restraints used for each H-bond identified were $r(\text{HN-O}) \leq 2.3$ Å, $r(\text{N-O}) = 2.5$ –3.3 Å. Initially 30 structures were computed starting from randomized coil structures using a dynamical simulated annealing protocol^{12,13} using the program X-PLOR. After two rounds of refinement in which the NOE restraints list was revised and added to, 16 structures had converged with good covalent geometry and van der Waals' contacts and no single violation greater than 0.5 Å. During the calculations the protein moves in a force field consisting of the sum of a number of terms. Bond (F_{bond}) and angle terms (F_{angle}) are used to maintain idealized covalent geometry. An improper dihedral term (F_{improper}) is incorporated to maintain the planarity and chirality

of certain covalent groups including the peptide bond which is assumed to be planar and universally *trans* in CD2 domain 1. Square well terms are used for experimental distance (F_{NOE}) and dihedral angle restraints (F_{cdihed}). A simple quartic repulsion term (F_{vdw}) is used to represent the van der Waals' component of the target function. The N-terminal region of the protein, comprising the Gly-Ser N-leader dipeptide and residues 1 and 2, was omitted from the calculations because of the absence of structurally significant NMR data. The final values of each term of the target function for the 16 structures were: $F_{\text{bond}} = 72 \pm 4$ kcal mol⁻¹; $F_{\text{angle}} = 1842 \pm 14$ kcal mol⁻¹; $F_{\text{improper}} = 100 \pm 4$ kcal mol⁻¹; $F_{\text{NOE}} = 168 \pm 18$ kcal mol⁻¹; $F_{\text{cdihed}} = 4.5 \pm 2$ kcal mol⁻¹; $F_{\text{vdw}} = 59 \pm 8$ kcal mol⁻¹; $F_{\text{total}} = 2,247 \pm 38$ kcal mol⁻¹. For each of these terms the final force constant used was identical to that given in ref. 12. The hard sphere van der Waals' radii were set to 0.8 of the standard values. Using the CHARMM²⁸ empirical energy function to measure a Lennard-Jones van der Waals' energy for each structure gave a value $E_{\text{L-J}} = -243 \pm 12$ kcal mol⁻¹, indicating good nonbonded contacts. Note that this Lennard-Jones term is not used as part of the target function in the structure calculations. The rms deviations from the experimental restraints and idealized covalent geometry are small: interproton distances (781) 0.085 ± 0.004 Å; experimental dihedral constraints (64) 1.047 ± 0.266 degree; bonds (1552) 0.010 ± 0.0003 Å; angles (2979) 2.078 ± 0.008 degree; improper dihedrals (634) 1.023 ± 0.018 degree. b, The backbone C α trace of the mean structure for CD2 domain 1 computed from the 16 best-fitted structures by coordinate averaging and energy minimization. The β strands are shown in blue, loops in red. c, The β sheets of CD2 superimposed with those of CD4 and REI V_L and NEW V_H. The structures were aligned with each other by matching 28 C α positions of the core residues of β strands B, C, D, E and F. d, A cutaway view of the superimposed CD2 structures showing the Ile-18 and Val-78 residues (with C α atoms coloured red) that are at the positions of the conserved IgSF disulphide bond in antibody V-domains. The separation of the C α atoms of these residues is the range observed for disulphide links in IgSF domains.

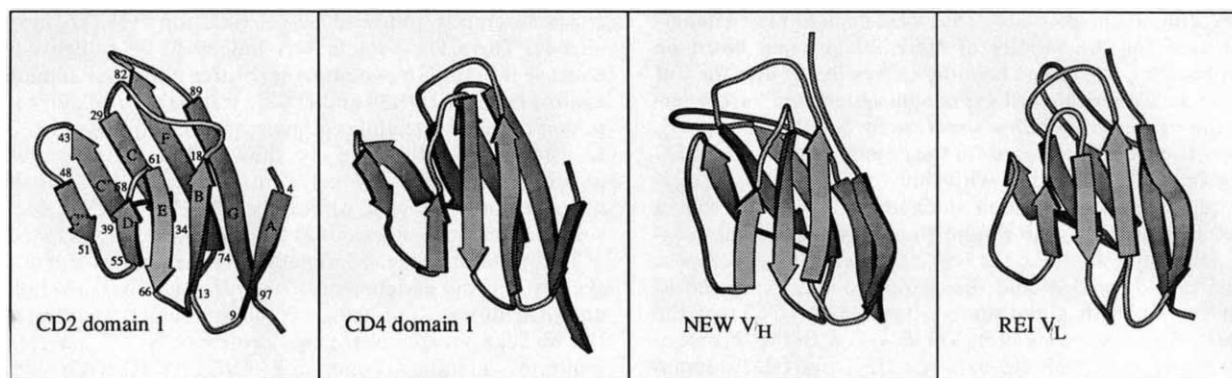


FIG. 3 Illustrations of the folds for CD2 domain 1, CD4 domain 1, immunoglobulin NEW V_H and immunoglobulin REI V_L . The coordinates for CD4 domain 1 were kindly provided by S. Harrison and those for the V domains were

from the Brookhaven database. The diagrams were computed using the MOLSCRIPT program²⁹.

calculations were done using the X-PLOR program based on dynamical simulated annealing^{12,13}. In total, 16 independent structures were obtained that satisfied the NOE distance restraints with no violation in any individual structure greater than 0.5 Å. The structures were based on 713 interproton NOE distance restraints, 64 ϕ dihedral angle restraints and 68 H-bond distance restraints. The 16 structures are shown in Fig. 2a, best-fitted on the backbone atoms of residues 3–98. A restrained minimized average structure was obtained by computing the mean coordinate positions of the 16 best-fitted structures followed by restrained minimization of the same target function as that used in the final stages of the structure calculation (Fig. 2b). The r.m.s. deviation of the 16 structures from the restrained minimized average structure is 0.98 ± 0.11 Å for backbone atoms and 1.59 ± 0.19 Å for all atoms in the well-defined portion of the structure spanning residues 3–98. This level of definition is sufficient to determine accurately the polypeptide fold of the protein and the conformation of some of the internal side chains.

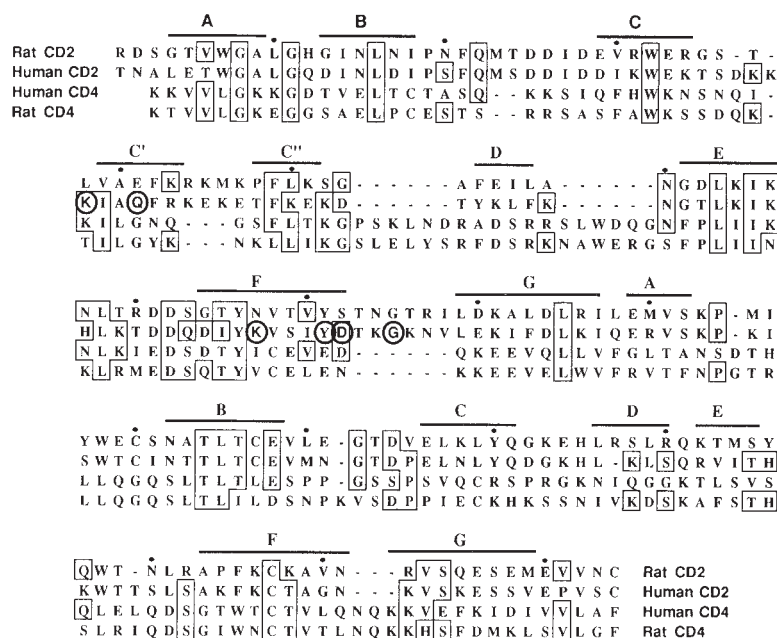
The calculated CD2 structures clearly show a fold like an immunoglobulin V-domain and in Fig. 2c the β strands for CD2 are overlaid with equivalent strands from CD4 (refs 14, 15) and the immunoglobulin V domains NEW V_H (ref. 16) and REI

V_L (ref. 17). It can be seen that the core strands are very similar with r.m.s. differences for the three-dimensional alignments being 0.92, 1.43 and 0.95 Å for CD2 versus CD4 and the V_H and V_L domains, respectively (calculations for 28 $C\alpha$ positions from β strands B, C, D, E and F).

Ile 18 and Val 78 are the residues of CD2 domain 1 corresponding to the conserved disulphide bond of IgSF domains⁵ and these residues are highlighted in Fig. 2d. In the 16 structures the $C\alpha$ atoms of these residues are separated by 7.0 ± 0.2 Å. In IgSF domains the $C\alpha$ - $C\alpha$ distance of the Cys residues forming the conserved disulphide bond is in the range 5.6–7.4 Å (ref. 18). In studies before the determination of the CD2 structure, cysteine residues were substituted by site-directed mutagenesis at Ile 18 and Val 78 and the mutant CD2 was expressed in a secreted two-domain form in Chinese hamster ovary cells. In this CD2 mutant a disulphide bond was formed (F. Gray, J.G.C., T. Willis, A.F.W. manuscript in preparation).

When the full CD2 domain is compared with CD4 domain 1 and the V-domains, differences in the regions connecting the β -strands are evident. Figure 3 shows comparisons of the folds for the four domains and the sequence comparisons for CD2 and CD4 are shown in Fig. 4. After the Ile-18 residue of β -strand

FIG. 4 Sequence alignments of CD2 and CD4 domains 1 and 2. Human and rat sequences are shown derived from the Swissprot database. The first three amino acids of human CD2 are not shown and the black dots above the rat CD2 sequence indicate each tenth residue of that sequence. The β -strand assignments have been made on the basis of the rat CD2 structure for the first domains and on the basis of the CD4 structure for the second domains. In the domain 1 sequences, the positions of CD4 β strands are also correctly identified by the bars but their extent is not exactly shown in all cases. Residues in human CD2 that are candidates for involvements in LFA-3 binding are circled and include human residues 43, 46, 82, 86, 87 and 90. These data are from mutational analysis¹⁹ (note residue 43 is equivalent to 48 in ref. 19 where the amino terminus was assigned five residues before the correct position). The criteria for considering a residue as a candidate for binding is that there should be a single mutation, not involving a Pro residue, that leads to the loss of LFA-3 binding and that the equivalent residue in the rat structure (38, 41, 77, 81, 82 and 85) is accessible to solvent. All of these residues except rat Ser 82 are exposed on the same β -sheet face.



B in CD2, there is a Pro residue and the CD2 chain folds immediately into a long, kinked loop connecting to β -strand C. In the V domains and CD4, β -strand B extends further and the loop joining strands B and C is not distorted. In CD2, β -strands D and E are truncated in comparison with the same regions in the other domains. This seems to be required to accommodate the kink in the loop between strands B and C since there are many internal side-chain contacts between the loops connecting strands B to C and D to E. These contacts plus the unusual structure for the B to C loop may explain the fact that sequence in this region is highly conserved between human and rat CD2. By contrast, in the same region of CD4 there is little conservation of sequence between the species (Fig. 4). In CD2 the loop between β -strand F and G is longer than in CD4 and is more similar to the same region in the V-domains where this loop constitutes hypervariable region 3.

The structure for CD4 (refs 14, 15) is unusual in comparison with immunoglobulin chains in that domain 2 follows on directly from domain 1 without any hinge residues such as are found between immunoglobulin V domains and the following C_L or C_{H1} domains. From Fig. 4 it is probable that the juxtaposition of domains in CD2 will be like that in CD4 as the distance of the β -strand G of domain 1 from β -strand B of domain 2 is one amino acid less in the CD2 sequence than in CD4.

The β -strand assignments for human CD2 domain 1 can be made by alignment with the newly solved rat CD2 domain 1 structure (Fig. 4). All the LFA-3 binding activity resides in domain 1 of CD2 (ref. 2) and mutagenesis of human CD2 has indicated regions of contact with LFA-3 (ref. 19). Residues which appear to be involved in this interaction are marked in Fig. 4. These are positioned in β -strand C' and in the region of β -strands F to G. This suggests that the face of the β -sheet C, C', F, G may be involved in the CD2:LFA-3 interaction. This can be further investigated by mutation analysis on the basis of the known three-dimensional structure.

The structure of CD2 domain 1 fits closely with that predicted from sequence on the basis of superfamily considerations with β -strands B, C, E and F that make up the core of the fold being exactly predicted^{5,7}. The proposal from secondary structure predictions^{2,6} is incorrect. CD2 domain 1 and CD4 domain 2 (refs 14, 15) were test cases for IgSF arguments because both sequences had unusual features which lead to contradictions between predictions from superfamily sequence patterns⁸ and *ab initio* computer calculations based on the primary sequences. The superfamily arguments were correct in both cases and this suggests that other IgSF domains predicted on the basis of conserved sequence patterns will also be correct²⁰.

The structure for CD2 domain 1 adds significantly to known variants of the immunoglobulin-fold as it is the first structure for a domain involved in cell adhesion and provides a prototype for IgSF domains that lack the conserved disulphide bond. Structures determined previously include immunoglobulin V and C domains²¹, MHC class 1 $\alpha 3$ domain²² and β -2 microglobulin^{22,23} and CD4 domains 1 and 2 (refs 14, 15). Also immunoglobulin-like folds have been seen in domains 1 and 2 of the PapD bacterial chaperone protein²⁴, but these should not be categorized as IgSF domains as they show no detectable sequence similarity to IgSF domains (ref. 25 and A.F.W., unpublished data). CD2 domain 1 typifies IgSF domains that lack the conserved disulphide bond. Such domains are found at a moderate frequency among domains of cell-surface molecules (for example LFA-3 domain 1, CEA domain 1, CD4 domain 3, PDGF receptor domain 4 (ref. 20)) but are the main type in the myosin-binding muscle proteins that are in the IgSF (ref. 26). □

3. Meuer, S. C. *et al. Cell* **36**, 897–906 (1984).
4. Beyers, A. D., Barclay, A. N., Law, D. A., He, Q. & Williams, A. F. *Immun. Rev.* **111**, 59–77 (1989).
5. Killeen, N., Moessner, R., Arvieux, J., Willis, A. & Williams, A. F. *EMBO J.* **7**, 3087–3091 (1988).
6. Clayton, L. K., Sayre, P. H., Novotny, J. & Reinherz, E. L. *Eur. J. Immun.* **17**, 1367–1370 (1987).
7. Williams, A. F., Barclay, A. N., Clark, S. J., Paterson, D. J. & Willis, A. C. *J. exp. Med.* **165**, 368–380 (1987).
8. Williams, A. F., Davis, S. J., He, Q. & Barclay, A. N. *Cold Spring Harb. Symp. quant. Biol.* **54**, 637–647 (1989).
9. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31–40 (1988).
10. Marion, D. *et al. Biochemistry* **28**, 6150–6156 (1989).
11. Driscoll, P. C., Clow, M., Marion, D., Wingfield, P. T. & Gronenborn, A. M. *Biochemistry* **29**, 3542–3556 (1991).
12. Clow, G. M. & Gronenborn, A. M. *CRC crit. Rev. Biochem. molec. Biol.* **24**, 479–564 (1989).
13. Brünger, A. T. *X-PLOR Manual, Version 2.1* (Yale University Press, 1990).
14. Wang, J. *et al. Nature* **348**, 411–418 (1990).
15. Ryu, S.-E. *et al. Nature* **348**, 419–426 (1990).
16. Saul, F. A., Amzel, M. & Poljak, R. J. *J. biol. Chem.* **253**, 585–597 (1978).
17. Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. *Biochemistry* **14**, 4943–4952 (1975).
18. Richardson, J. S. *Adv. Protein Chem.* **34**, 167–339 (1981).
19. Peterson, A. & Seed, B. *Nature* **329**, 842–846 (1987).
20. Williams, A. F. & Barclay, A. N. *A. Rev. Immun.* **6**, 381–405 (1988).
21. Amzel, L. M. & Poljak, R. J. *A. Rev. Biochem.* **48**, 961–967 (1979).
22. Bjorkman, P. J. *et al. Nature* **329**, 506–512 (1987).
23. Becker, J. W. & Reeke, G. N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4225–4229 (1985).
24. Holmgren, A. & Branden, C.-I. *Nature* **342**, 248–251 (1989).
25. Freeman, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 8810–8814 (1990).
26. Labiet, S. *et al. Nature* **345**, 273–276 (1990).
27. Kay, L. E. & Bax, A. *J. Magn. Reson.* **86**, 110–126 (1990).
28. Brooks, B. R. *J. comput. Chem.* **4**, 187–217 (1983).
29. Kraulis, P. J. *J. appl. Crystallogr.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Boyd, T. Harvey and N. Soffe from the Oxford Department of Biochemistry for helpful discussions and S. Harrison, Harvard University for supplying the CD4 coordinates. I.D.C. is a member of the Oxford Centre for Molecular Sciences supported by the SERC and MRC. P.C.D. is supported by a Royal Society University Research Fellowship and J.G.C. by a Commonwealth Overseas Studentship.

Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens

Suzanne B. Hartley^{*†}, Jeffrey Crosbie^{*},
Robert Brink^{*}, Aaron B. Kantor[‡],
Antony Basten^{*} & Christopher C. Goodnow[†]

^{*} Centenary Institute of Cancer Medicine and Cell Biology, University of Sydney, New South Wales 2006, Australia

[†] Howard Hughes Medical Institute, and Department of Microbiology and Immunology, and [‡]Department of Genetics, Beckman Center for Molecular and Genetic Medicine, Stanford University, Stanford, California 94305-5428, USA

THE long-standing hypothesis^{1,2} that tolerance to self antigens is mediated by either elimination^{3–8} or functional inactivation (anergy; refs 9–11) of self-reactive lymphocytes is now accepted, but little is known about the factors responsible for initiating one process rather than the other. In the B-cell lineage, tolerant self-reactive cells persist in the peripheral lymphoid organs of transgenic mice expressing lysozyme and anti-lysozyme immunoglobulin genes⁹, but are eliminated in similar transgenic mice expressing anti-major histocompatibility complex immunoglobulin genes⁸. By modifying the structure of the lysozyme transgene and the isotype of the anti-lysozyme immunoglobulin genes, we demonstrate here that induction of anergy or deletion is not due to differences in antibody affinity or isotype, but to recognition of monomeric or oligomeric soluble antigen versus highly multivalent membrane-bound antigen. Our findings indicate that the degree of receptor crosslinking can have qualitatively distinct signalling consequences for lymphocyte development.

The soluble lysozyme gene construct used earlier⁹ was modified to encode an integral membrane protein by inserting a complementary DNA fragment encoding the extracellular spacer sequence, transmembrane segment and cytoplasmic tail

Received 4 July; accepted 23 August 1991.

1. Springer, T. A. *Nature* **346**, 425–434 (1990).
2. Recny, M. A., Neidhardt, E. A., Sayre, P. H., Ciardelli, T. L. & Reinherz, E. L. *J. biol. Chem.* **265**, 8542–8549 (1990).

of the H-2K^b class I major histocompatibility complex (MHC) molecule (Fig. 1a). In addition, the metallothionein promoter that we used previously⁹ was replaced with the more ubiquitous promoter from the H-2K^b gene¹² to ensure that the membrane lysozyme would be expressed on cells in the bone marrow where clonal deletion is thought to take place. Transfection of the resulting construct confirmed that lysozyme was expressed on the cell surface (Fig. 1a). Transgenic mice carrying the membrane lysozyme gene construct were produced using fertilized eggs from inbred C57BL/6 mice and one line, KLK3, was initially selected for further analysis. Fluorescence activated cell sorting (FACS) of spleen, lymph node, bone marrow and thymus from KLK3 transgenic offspring revealed membrane lysozyme expression on many of the cells in these organs (Fig. 1b).

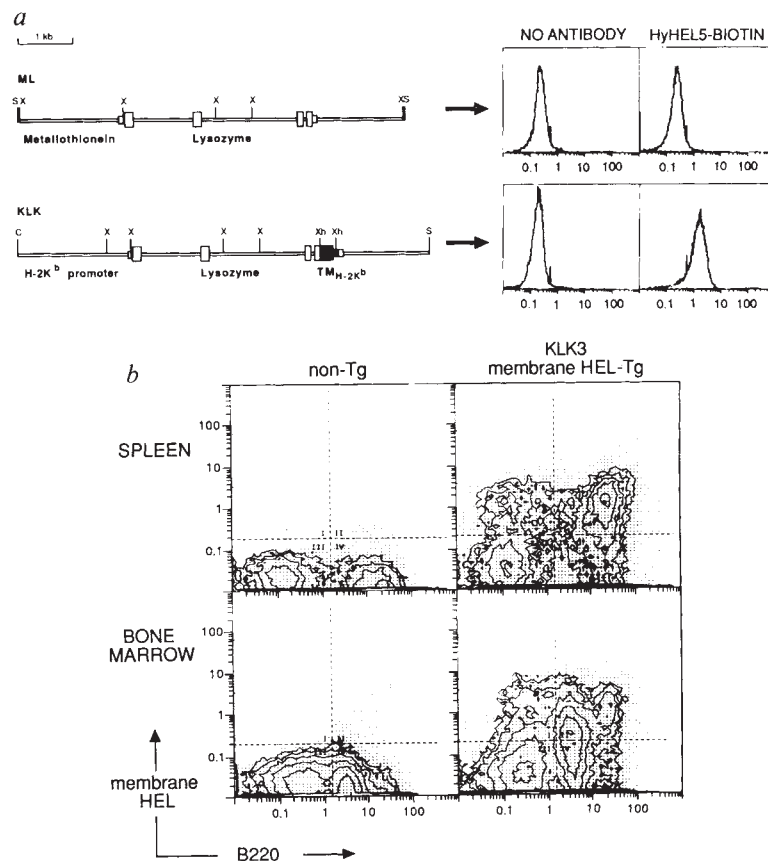
Double-transgenic mice were generated by mating membrane-lysozyme transgenic or soluble-lysozyme transgenic mice with immunoglobulin (Ig)-transgenic mice carrying rearranged anti-lysozyme heavy- and light-chain genes. As the deleted B cells in the anti-MHC transgenic mice expressed only μ -heavy chains⁸, the possibility that elimination or anergy was the consequence of the absence or presence of membrane IgD was investigated by producing Ig-transgenic mice that carried heavy-chain genes containing either the entire μ and δ -constant-region locus (MD4; ref. 13) or a truncated heavy-chain gene containing the μ -constant-region locus but lacking the exons for IgD (MM4; ref. 13). Both types of membrane lysozyme double-transgenic mice, like those expressing soluble lysozyme⁹, had undetectable levels of serum anti-lysozyme antibody, indicating

that tolerance to lysozyme is maintained in the double-transgenic animals. By contrast, FACS analysis of spleen cells (Fig. 2a, b) and lymph node cells (not shown) shows that whereas lysozyme-binding cells persist in the soluble-lysozyme double-transgenic mice, these cells are undetectable in the membrane-lysozyme double-transgenic animals (Fig. 2a, top). Moreover, the number of mature B cells, as estimated by high-expression of the B220 antigen (B220^{high}; refs 14, 15), was reduced 5–20 fold in the spleen, lymph node and bone marrow of membrane-lysozyme double-transgenic mice (Fig. 2), irrespective of whether the animals carried μ or $\mu + \delta$ heavy-chain transgenes. The residual mature B cells seem to be comprised of the normally minor subpopulations of cells expressing either endogenous b-allotype heavy chains (not shown) or transgenic a-allotype heavy chains that bind very little or undetectable amounts of lysozyme (Fig. 2a). The lack of lysozyme binding to these cells may be due to expression of endogenous light chains which generally preclude high-affinity lysozyme binding⁹.

Depletion of mature lysozyme-reactive B cells was confirmed in radiation chimaeras constructed by reconstituting membrane-lysozyme transgenic animals with MD4 Ig-transgenic bone marrow. The advantage of these chimaeras was that the sensitive hen egg lysozyme (HEL)/HyHEL5-biotin sandwich stain⁹ could be used to detect any residual lysozyme-specific B cells without the complication of membrane-lysozyme expression on the transgenic B cells which occurred in the membrane-lysozyme double-transgenic mice. As shown in Fig. 3a, no B220^{high} B cells bearing lysozyme-binding immunoglobulin were detected in the

FIG. 1 Production of transgenic mice expressing membrane-bound lysozyme. **a**, Map of the original *ML* gene construct encoding soluble lysozyme and the modified *KLK* construct encoding membrane-bound lysozyme. The metallothionein and H-2K^b promoters are indicated, 5' and 3' untranslated regions are represented by open boxes and the coding regions by wider boxes. The introduced transmembrane sequence derived from the H-2K^b cDNA is shown in black. Restriction sites indicated are: S, *Sall*; X, *Xba*I; Xh, *Xho*I; C, *Cla*I. Shown in the adjacent panel are one-colour FACS analyses of Sp2/0 cells transfected with each construct after staining the cells with a biotinylated anti-lysozyme monoclonal antibody, HyHEL5 (ref. 30) to detect surface lysozyme. FACS histograms are displayed as fluorescence emission on the x axis versus the number of cells. Equivalent staining was obtained with antibodies to three different conformational epitopes of HEL (ref. 30), indicating that the folding of the extracellular lysozyme domain was indistinguishable from the native molecule. **b**, Two-colour FACS analyses of spleen and bone marrow cells from non-transgenic (non-Tg) and KLK3 membrane-bound-lysozyme transgenic (membrane HEL-Tg) mice, stained for membrane lysozyme and for the B220 antigen³¹. Green and yellow fluorescence is displayed in arbitrary relative fluorescence units, with dots representing one or more cells and contour intervals set at 5, 10, 20, 40, 80 and 160 cells per square of a 4 × 4 channel grid.

METHODS. The *KLK* membrane-lysozyme gene construct was prepared from a modified version of the original *ML* construct, pHEL/CYT (W. Ho and M. M. Davis, manuscript in preparation), which contained a unique *Xho*I site in place of the lysozyme stop codon. The H-2K^b cDNA transmembrane sequence was inserted into the *Xho*I site by first digesting the H-2K^b cDNA (ref. 32; gift from J. F. A. P. Miller) with *Rsa*I and *Pvu*II and then adding *Xho*I linkers. The H-2K^b promoter was derived as a 2.0-kb *Nru*I–*Hind*III fragment from the H-2K^b gene³³ (gift from J. F. A. P. Miller), subcloned into pBluescript, removed by *Cla*I and partial *Xba*I digestion, and inserted in place of the metallothionein promoter. Stable transfection of Sp2/0 cells was as described³⁴. Representative transfectants were stained with biotinylated anti-lysozyme antibodies, HyHEL5, HyHEL9 or HyHEL10 (ref. 30), followed by streptavidin-phycoerythrin (Caltag), and analysed on a FACS 440. For oocyte microinjection, the *KLK* construct was excised from the vector with *Cla*I and *Sall*,



purified, and microinjected into C57BL/6 eggs as before⁹. Twelve transgenic founders were obtained from 86 offspring, and preliminary analyses indicated that most of the founders expressed the transgene. Spleen and bone marrow cells from KLK3 transgenic offspring were stained with HyHEL5-biotin/streptavidin-phycoerythrin and fluorescein-RA3-6B2, and analysed as described^{9,35}.

FIG. 2 Elimination of mature lysozyme-reactive B cells and persistence of immature B cells in membrane-lysozyme double-transgenic mice. **a**, Transgenic mice expressing anti-lysozyme IgM and IgD. Spleen cells were obtained from non-transgenic (non-Tg), Ig-transgenic (Ig-Tg) and membrane-lysozyme double-transgenic (Dbl-Tg) littermates and from soluble-lysozyme double-transgenic mice of similar age. Cells were stained for B220 and counterstained for lysozyme binding (HEL-biotin, top row), or with a sandwich technique capable of detecting cell-bound lysozyme as well as binding of exogenous lysozyme (HEL/HyHEL5-biotin, bottom row). Middle panels show staining for transgene-encoded α -allotype IgM (IgM^a) or IgD (IgD^a). Windows denote B cells expressing low or high levels of B220 and were determined by parallel stains of immature (B220^{low}) and mature (B220^{high}) B cells in the bone marrow^{14,15}. The number of mature (B220^{high}) B cells staining positive for each of the fluorescent labels is shown as a percentage of splenic lymphocytes. **b**, Equivalent transgenic mice expressing IgM only, with spleen cells stained as in the bottom row of **a**. **c**, Bone marrow from the same mice, showing persistence of immature B cells expressing transgene-encoded IgM (IgM^a). The number of B220^{high} and B220^{low} IgM^a-positive cells is expressed as a percentage of bone marrow lymphocytes. **d**, Relative frequency of total B220^{low} or B220^{high} cells in different lymphoid compartments of IgM + D Ig-transgenic mice and membrane-lysozyme double-transgenic littermates. Dots show the percentage of B220-bearing lymphocytes in individual mice from three pairs of littermates. **METHODS**. Offspring from matings of MD4 $\times$ KLK3 or MD4 $\times$ ML5 (refs 9, 35) were analysed at 8–12 weeks of age in **a** and from MM4 $\times$ KLK3 or MM4 $\times$ ML5 at 8 weeks of age in **b**. Bone marrow, spleen and lymph node cells were stained and FACS-analysed as described^{9,35}. The absolute numbers of B220^{high} B cells in the spleen were ($\times 10^{-7}$): **a**, Ig-Tg, 4.1; soluble-HEL Dbl-Tg, 2.9; membrane-HEL Dbl-Tg, 0.6; **b**, Ig-Tg, 2.7; soluble-HEL Dbl-Tg, 1.7; membrane-HEL Dbl-Tg, 0.2.

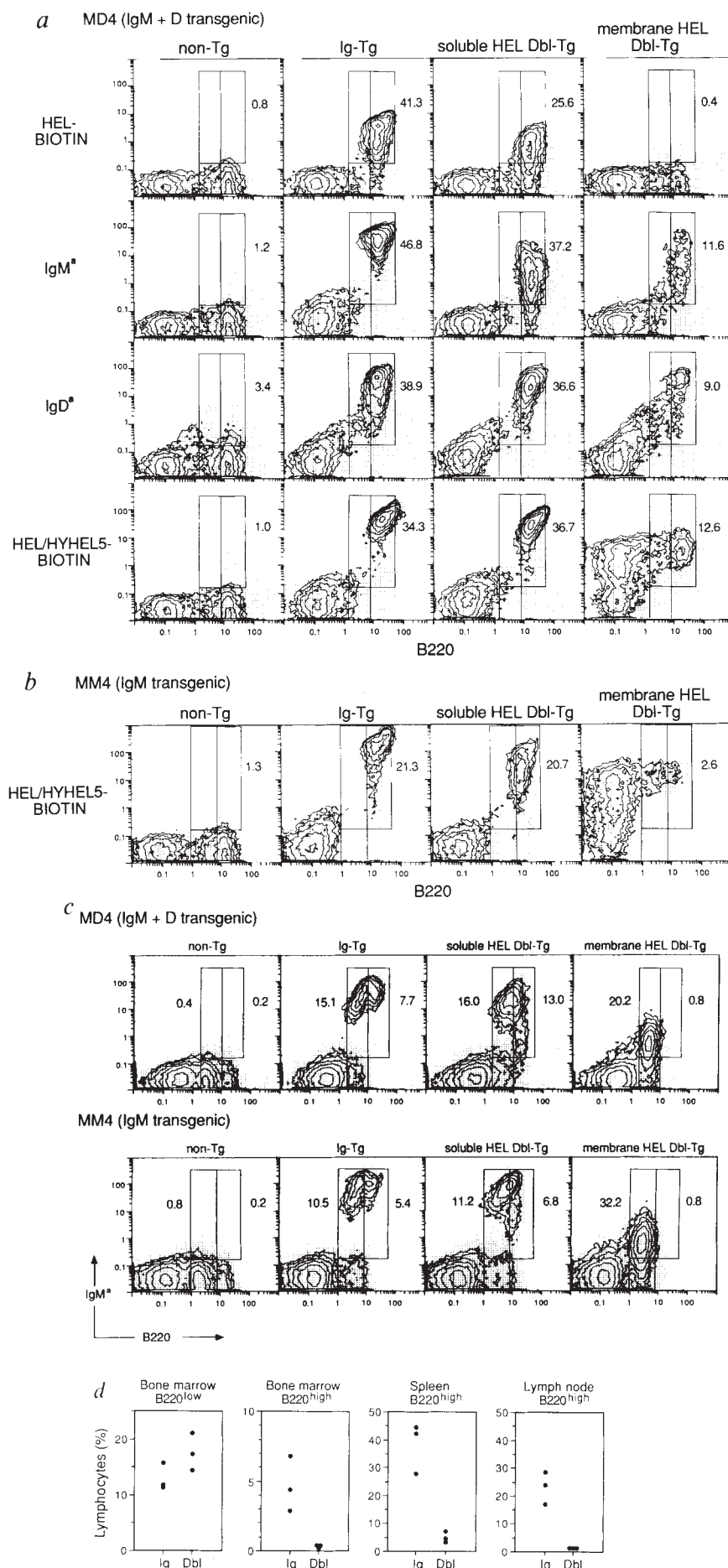
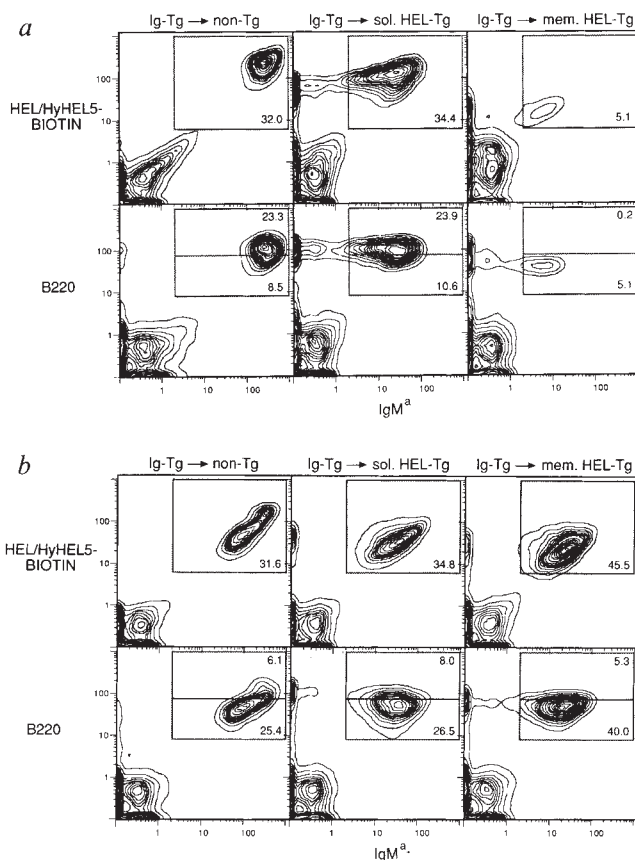


FIG. 3 Elimination of mature lysozyme-reactive B cells from the spleen of membrane-lysozyme-expressing radiation chimaeras. Pairs of lethally irradiated membrane-lysozyme transgenic recipients, or non-transgenic and soluble-lysozyme transgenic controls, were reconstituted with bone marrow cells from an anti-lysozyme Ig-transgenic (IgM^a/IgD^b) mouse. Spleen and bone marrow cells from the reconstituted recipients were stained for lysozyme binding (HEL/HyHEL5), transgene-encoded IgM^a, B220 and endogenous IgM^b/IgD^b, and analyzed by four-colour FACS. In the upper row of *a* (spleen), staining for IgM^a and for lysozyme-binding shows that few lysozyme-binding B cells remain in the membrane-lysozyme transgenic recipients (the percentage of splenic lymphocytes falling within the window is shown), but that the remaining cells nevertheless bind lysozyme in proportion to their membrane IgM^a density, indicating that they also express transgene-encoded light chains. In the lower row, staining of the same cells for B220 demonstrates that all of the lysozyme-reactive B cells remaining are immature (B220^{low}) cells, whereas most of the B cells present in the control chimaeras are mature (B220^{high}). In the bone marrow (*b*), equal or increased numbers of HEL-binding B220^{low} B cells are present in the membrane-lysozyme transgenic recipients compared with the non-transgenic and soluble-lysozyme transgenic recipient controls. Equivalent results were obtained from both animals in each pair.

METHODS. Recipients were exposed to two doses of 450 rad X-irradiation three hours apart, and reconstituted 6 h later by intravenous injection of 2.5×10^6 bone marrow cells from an MD4 Ig-transgenic mouse. Four weeks later, spleen and bone marrow cell suspensions were prepared and stained as described^{9,35}, except that the anti-IgM^a monoclonal antibody, DS-1 (ref. 36), was directly conjugated to allophycocyanin, and the RA3-6B2 staining was developed with anti-rat F(ab')₂-Texas red (Caltag). Fifty thousand spleen cells and 10^5 bone marrow cells were analysed on a modified FACS II in the Beckman Center Shared FACS Facility, using the FACS/DESK analysis program³⁷, with contour levels set at a 5% probability. The total number of lysozyme-binding, IgM^a-positive cells in the spleen were ($\times 10^{-7}$): Ig-Tg into non-Tg, 1.9; Ig-Tg into soluble-HEL-Tg, 1.5; and Ig-Tg into membrane-HEL-Tg, 0.2.



spleens of membrane lysozyme-expressing recipients, whereas large numbers of these cells were still present in reconstituted soluble-lysozyme transgenic recipients and in nontransgenic recipient controls. Equally efficient elimination of lysozyme-reactive B cells was also observed in a second set of radiation chimaeras constructed by reconstituting nontransgenic recipients with a 9:1 mixture of immunoglobulin transgenic and membrane-lysozyme transgenic bone marrow cells (not shown). Taken together, the findings from the two sets of chimaeras indicate that elimination of self-reactive B cells could be triggered by membrane lysozyme expressed either on radio-resistant tissues or on a trace number of haemopoietic cells, and was not due to expression of membrane lysozyme by the B cell itself or to retention of B cells on lysozyme-expressing vascular endothelium.

In contrast to the absence of mature lysozyme-reactive B cells in the membrane-lysozyme double-transgenic mice and chimaeras, the number of immature (B220^{low}) surface Ig-positive B cells in the bone marrow of these animals was not reduced, and in fact appeared to be slightly increased (Figs 2c, d and 3b). The forward and side-scatter profiles of these immature B cells were unaltered (not shown) and revealed no evidence of the characteristic changes detected in cells undergoing programmed cell death¹⁶. Nevertheless, the immature lysozyme-reactive B cells showed clear evidence of having encountered self antigen, as they exhibited a marked decrease in membrane IgM expression relative to the level expressed on cells from immunoglobulin-transgenic littermate controls. The presence of large numbers of immature lysozyme-reactive B cells in mice expressing membrane lysozyme contrasts with the low numbers of analogous self-reactive B cells reported in the anti-MHC model¹⁷. The latter investigators however did observe a persistent population of B220^{low/intermediate} pre-B cells, most of which seem to be immature B cells expressing low levels of surface immunoglobulin (ref. 17, and Figs 2 and 3). The persistence of autoreactive cells within the immature B-cell pool presents an

intriguing contrast to that described in the T-cell lineage, where there is a drastic loss of immature thymocytes^{6-7,18}. The difference could reflect a more prolonged process of receptor-induced apoptosis in the B-cell lineage, or may indicate that B-cell elimination instead occurs by default, following a failure or arrest of maturation^{2,19,20}.

Our finding of elimination of self-reactive B cells recognizing cell-surface lysozyme resolves the disparity in earlier studies of tolerance in immunoglobulin-transgenic mice^{8,9} by showing induction of elimination or anergy in immature B cells to be dependent on the degree of antigen-receptor crosslinking. The fact that soluble lysozyme induced anergy in immature B cells without aborting their maturation strongly suggests that immature B cells have two signalling thresholds, one induced by extensive receptor crosslinking and the other by inefficient crosslinking or possibly by receptor occupancy alone. This conclusion reinforces the long-held view that signalling through the B-cell antigen receptor is not an all-or-none event, but that quantitative differences in the degree of receptor crosslinking may bring about qualitatively different cellular responses^{2,21-26}. As multiple signalling pathways have now been linked to the B-cell antigen receptor²³⁻²⁹, it will be interesting to determine whether the degree of receptor crosslinking affects quantitative or qualitative differences in these second messengers. Moreover, the tendency for membrane antigen-reactive B cells to be purged from the mature B-cell repertoire raises questions concerning the origins of autoantibodies to cell-surface structures in diseases such as myasthenia gravis, autoimmune haemolytic anaemia and autoimmune thrombocytopenic purpura. □

Received 15 March; accepted 5 September 1991.

1. Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (Vanderbilt University Press, Nashville, 1959).
2. Nossal, G. J. V. *A. Rev. Immun.* **1**, 33-62 (1983).
3. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
4. Kappler, J. W., Staerz, U., White, J. & Marrack, P. *C. Nature* **332**, 35-40 (1988).

5. MacDonald, H. R. *et al.* *Nature* **332**, 40–45 (1988).
6. Kisielow, P., Blüthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742–746 (1988).
7. Sha, W. C. *et al.* *Nature* **336**, 73–76 (1988).
8. Nemazee, D. A. & Bürki, K. *Nature* **337**, 562–566 (1989).
9. Goodnow, C. C. *et al.* *Nature* **334**, 676–682 (1988).
10. Rammensee, H.-G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541–544 (1989).
11. Ramsdell, F., Lantz, T. & Fowlkes, B. J. *Science* **246**, 1038–1041 (1989).
12. Morello, D., Moore, G., Salmon, A. M., Yaniv, M. & Babinet, C. *EMBO J.* **5**, 1877–1883 (1986).
13. Goodnow, C. C. *et al.* in *The Immune Response to Structurally Defined Proteins: the Lysozyme Model* (eds Smith-Gill, S. & Sercarz, E.) 389–402 (Adenine, New York, 1989).
14. Forster, I., Vieira, P. & Rajewsky, K. *Int. Immun.* **1**, 321–331 (1989).
15. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. *J. exp. Med.* **173**, 1213–1225 (1991).
16. Wyllie, A. H., Kerr, J. F. R. & Currie, A. R. *Int. Rev. Cytol.* **68**, 251–306 (1980).
17. Nemazee, D. & Bürki, K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8039–8043 (1989).
18. Murphy, K. M., Heimberger, A. B. & Loh, D. Y. *Science* **250**, 1720–1723 (1990).
19. Sidman, C. L. & Unanue, E. R. *Nature* **257**, 149–151 (1975).
20. Raff, M. C. *et al.* *J. exp. Med.* **142**, 1052–1064 (1975).
21. Klaus, G. G. B., Howard, J. G. & Feldmann, M. *Br. med. Bull.* **32**, 141–146 (1976).
22. Dintzis, R. Z., Middleton, M. H. & Dintzis, H. M. *J. Immun.* **131**, 2196–2203 (1983).
23. DeFranco, A. L. *Rev. Cell Biol.* **3**, 143–178 (1987).
24. Cambier, J. C. & Ransom, J. T. *A. Rev. Immun.* **5**, 175–199 (1987).
25. Harnett, M. M. & Klaus, G. G. B. *Immun. Today* **9**, 315–320 (1988).
26. Gold, M. R., Law, D. A. & DeFranco, A. L. *Nature* **345**, 810–813 (1990).
27. Campbell, M. A. & Sefton, B. M. *EMBO J.* **9**, 2125–2131 (1990).
28. Brunswick, M., Samelson, L. E. & Mond, J. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1311–1314 (1991).
29. Reth, M. *et al.* *Immun. Today* **12**, 196–201 (1991).
30. Smith-Gill, S. J., Mainhart, C. R., Lavoie, T. B., Rudikoff, S. & Potter, M. *J. Immun.* **132**, 963–967 (1984).
31. Coffman, R. L. *Immunol. Rev.* **62**, 5–23 (1982).
32. Schönrich, G. *et al.* *Cell* **65**, 293–304 (1991).
33. Weiss, E. *et al.* *EMBO J.* **2**, 453–462 (1983).
34. Gascoigne, N. R. J., Goodnow, C. C., Dudzik, K. I., Oi, V. T. & Davis, M. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2936–2940 (1987).
35. Goodnow, C. C., Crosbie, J., Jorgensen, H., Brink, R. A. & Basten, A. *Nature* **342**, 385–391 (1989).
36. Stall, A. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 3546–3550 (1988).
37. Parks, D. R., Lanier, L. L. & Herzenberg, L. A. in *The Handbook of Experimental Immunology* Vol. 1 (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 29.1–29.21 (Blackwell, Edinburgh, 1986).

ACKNOWLEDGEMENTS. We thank M. M. Davis and W. Ho for pHEL/CYT; J.F.A.P. Miller, B. Arnold and E. Weiss for the H-2K^b cDNA and genomic clones; P. Gregory and H. Jorgensen for advice and assistance; C. Sagerstrom and C. Guidos for help with the chimaera analyses; and A. Cheung and J. Ferraz for animal husbandry. This research was funded by grants from the NH&MRC and by the HHMI.

Hypervariable C-terminal domain of rab proteins acts as a targeting signal

Philippe Chavrier, Jean-Pierre Gorvel, Ernst Stelzer, Kai Simons, Jean Gruenberg & Marino Zerial*

European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, Germany

MAMMALIAN cells express many ras-like low molecular mass GTP-binding proteins (rab proteins)^{1–6} that are highly homologous to the Ypt1 and Sec4 proteins involved in controlling secretion in yeast^{7–9}. Owing to their structural similarity and to their variety, rab proteins have been postulated to act as specific regulators of membrane traffic in exocytosis and endocytosis¹⁰, and rab5 has been shown to be involved in early endosome fusion *in vitro*¹¹. In agreement with their postulated functions, all rab proteins studied so far have been found in distinct subcompartments along the exocytic or endocytic pathways^{5,11–13}. To define the region mediating their specific localization, we transiently expressed rab2, rab5 and rab7 hybrid proteins in BHK cells, and determined their intracellular localization by immunofluorescence confocal microscopy and subcellular fractionation. Here we present evidence that the highly variable C-terminal domain contains structural elements necessary for the association of rab proteins with their specific target membranes in the endocytic pathway.

Ras and ras-related proteins share four highly conserved regions which cooperate to form the GTP-binding site. They also have one or two cysteine residues near their C termini^{1–8} (see Fig. 1a) which are required for membrane association^{5,14–19}. The C-terminal domain (downstream from the fourth conserved region among ras-related proteins⁶; Fig. 1a) represents a candidate signal for targeting rab proteins to their corresponding acceptor membranes. It is located adjacent to the cysteine motif and must therefore lie close to the membrane and is highly variable in both sequence and length.

To test whether the C terminus could act as a targeting signal, we expressed chimaeric proteins made up of portions of rab5 and rab7 (Fig. 1b) in BHK cells using the T7 RNA polymerase recombinant vaccinia virus system^{20,5,11}. The distribution of the chimaeric proteins was analysed by immunofluorescence confocal microscopy and subcellular fractionation. Conditions were first established by cotransfecting BHK cells with wild-type rab5 or rab7 constructs and a plasmid encoding the human transferrin receptor (hTR)²¹, a marker for plasma membrane and early endosomes^{22,23}. We had previously shown that rab5 is associated

with the plasma membrane and early endosomes, whereas rab7 is restricted to late endosomes^{5,11}.

Affinity-purified rabbit antisera raised against peptides derived from the C-terminal variable regions of rab5 and rab7 proteins⁵ (Fig. 1b) were used at concentrations selected to detect only the overexpressed but not the endogenous rab proteins, and were visualized using Texas Red-conjugated goat anti-rabbit antibodies. A monoclonal antibody, B3-25, specific for the hTR was identified with FITC-labelled goat anti-mouse antibodies. Colocalization or superimposition of the two chromophores was revealed by the yellow colour resulting from their overlapping emissions. Most of the transiently expressed rab5 and hTR localized together as expected (Fig. 2a, b, c). In contrast, segregation of the two chromophores was clearly observed in cells expressing rab7 and hTR: rab7 was associated with structures located perinuclear to those containing hTR (Fig. 2d–f), consistent with its localization to late endosomes⁵. The superimposition of the two chromophores in the perinuclear region may account for the accumulation of rab7 label in this area.

Further evidence for the localization of rab5 and rab7 was obtained by subcellular fractionation. Fractions enriched in early and late endosomes were separated by flotation using a sucrose-D₂O gradient¹¹. Previously, we have shown that the early endosome fraction was enriched in rab5, whereas the late endosome fraction was enriched in rab7 and the cation-independent mannose-6-phosphate receptor. The overexpressed rab5 and rab7 proteins were detected enriched in the early and late endosome fractions (Fig. 3a), respectively, thus confirming our previous observation that high expression levels do not lead to mislocalization^{5,11}.

We then tested whether exchange of the C-terminal cysteine motif could change the localization of these proteins. When we exchanged the CCSN motif of rab5 with the CSC sequence of rab7, yielding rab5CSC, the hybrid protein displayed an immunofluorescence staining pattern very similar to that of wild-type rab5 and overlapping with hTR (not shown). Rab5CSC, like wild-type rab5, was detected in the early endosome fraction (Fig. 3a). This indicates that specific targeting is not mediated solely by the cysteine motif. We then investigated whether sequences upstream of the cysteine motif of rab7 were involved in targeting. We replaced the last 8 (Nrab5₂₀₆Crab7₂₀₀) and 13 (Nrab5₂₀₁Crab7₁₉₅) C-terminal residues of rab5 with the corresponding sequences from rab7. Both hybrid proteins were detected in the early endosome fraction. Double immunofluorescence confocal microscopy confirmed that the staining pattern was very similar to that of hTR (not shown). In contrast, an N-terminal rab5 reporter fragment of 181 residues fused to the C-terminal 34 amino acids of the rab7 protein (Nrab5₁₈₁-Crab7₁₇₄) displayed a similar localization to rab7 in the perinuclear region (compare Fig. 2d, e, f and g, h, i). This hybrid protein was detected in the late endosome fraction using

* To whom correspondence should be addressed.

FIG. 1 a, Scheme of a typical ras-like low molecular mass GTP-binding protein showing the variable lengths of the N and C termini, the conserved regions participating in the formation of the GTP-binding site (stippled boxes) including the highly conserved DTAGQE motif (black box), the 'effector region' (striped box) and the C-terminal cysteine motifs⁶. **b**, Diagrams of rab5, rab7 and rab2 wild-type and hybrid proteins and their localization to early and late endosome fractions. Alignment was done on the WDTAGQE motif (single-letter amino-acid notation) shared by all rab proteins (black box)⁶. Numbers, amino-acid residues of wild-type and hybrid proteins; bars, positions corresponding to the synthetic peptides recognized by the antibodies. **METHODS**. All wild-type and hybrid complementary DNAs were inserted in plasmid pGEM1 under the control of the T7 RNA polymerase promoter. Plasmids pGEM1-rab2, -rab5 and -rab7 have been described previously⁵. Plasmid pGEM1-rab5CSC codes for the rab5 protein having the C-terminal CCSN sequence replaced by CSC from rab7. Plasmids pGEM1-Nrab5₂₀₆Crab7₂₀₀ and pGEM1-Nrab5₂₀₁Crab7₁₉₅ code for hybrid proteins consisting of the N-terminal 206 residues of rab5 fused to the C-terminal eight residues of rab7 and the N-terminal 201 residues of rab5 fused to the C-terminal 13 residues of rab7, respectively. Plasmid pGEM1-Nrab5₁₈₁Crab7₁₇₄ encodes a fusion protein between the 181 N-terminal residues of rab5 and the 34 C-terminal residues of rab7. Plasmid pGEM1-rab5Δ₂₀₁₋₂₁₁ encodes a rab5 protein having a deletion of 10 amino acids between position 201 and 211. pGEM1-Nrab2₁₆₆Crab5₁₈₁ encodes a rab2 N-terminal fragment of 166 amino acids fused to the 35 C-terminal residues of rab5 (position 181-215). Plasmid pGEM1-Nrab2₁₆₆Crab7₁₇₃ encodes a fusion protein between the 166 N-terminal residues of rab2 and the 35 C-terminal residues of rab7. All hybrid inserts were obtained using specific oligonucleotides in a polymerase chain reaction (PCR)-based approach²⁵ and their open reading frames were completely checked by nucleotide sequencing. All hybrid proteins bound [³²P]GTP when immobilized on nitrocellulose filter (not shown).

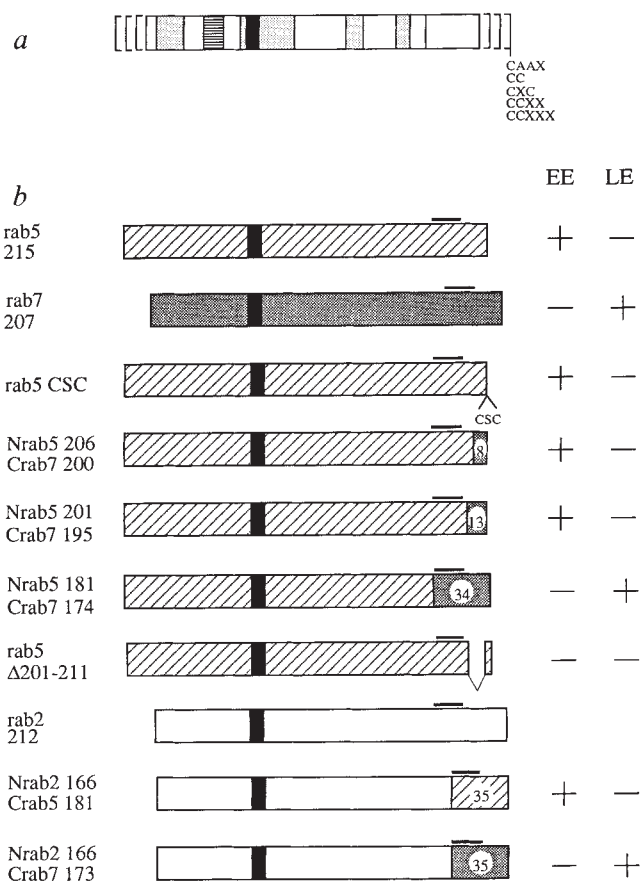
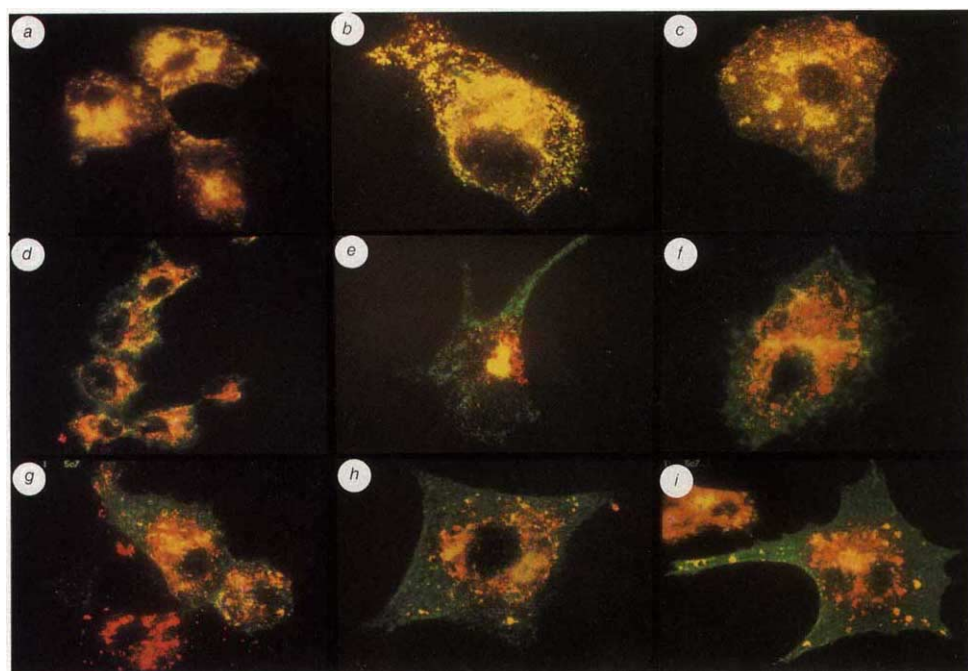


FIG. 2 Immunofluorescence localization of rab wild type, hybrid rab proteins and human transferrin receptor (hTR) coexpressed in BHK cells. **a, b, c**, rab5, hTR; **d, e, f**, rab7, hTR; **g, h, i**, Nrab5₁₈₁Crab7₁₇₄, hTR.

METHODS. Double immunofluorescence staining of BHK-transfected cells was carried out as previously described⁵ using affinity-purified rabbit anti-rab proteins antibodies (as indicated in Fig. 1). Anti-rab antibodies were visualized with Texas Red conjugated goat anti-rabbit antibodies. Cells were permeabilized with 0.5% saponin before fixation to wash out excess cytosolic overexpressed rab proteins. Only transfected cells are detected at the concentrations of antibodies used. The hTR protein was detected using the monoclonal antibody B3-25 (Boehringer) and goat anti-mouse antibodies conjugated with FITC. Most of the transfected cells expressed both the rab and the hTR proteins. Coverslips were viewed with the confocal microscope developed at EMBL. The wavelength of 514 nm was used to excite the Texas Red and FITC chromophores. Images of one optical section (0.4 μm on the z axis) taken from the base of the cells were photographed on a Fujichrome 100D film with a Polaroid freeze frame directly from monitor.



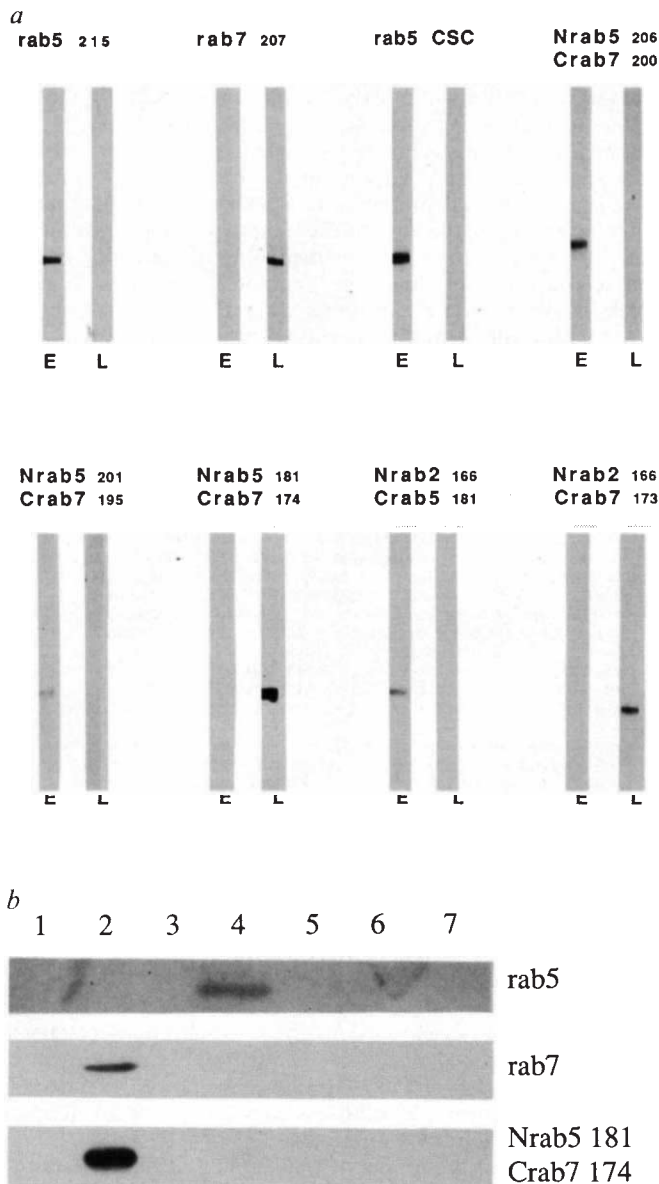


FIG. 3 Immunoblot analysis of rab wild type and hybrid proteins in (a) early (E) or late (L) endosome-enriched fractions and (b) in fractions from the flotation gradient. c, Distribution of rab5 and rab5 Δ 201-211 in high-speed pellet (P) and supernatant (S) of transfected BHK cells.

METHODS. BHK cells were infected with the T7 RNA polymerase recombinant vaccinia virus and transfected with the different plasmids as described previously⁵. Preparation and analysis of the fractions from the flotation gradient were carried out according to Gorvel *et al.*¹¹ After centrifugation the gradient was manually divided into seven fractions: fraction 2 was late endosome-enriched and corresponded to the uppermost portion of the 10% sucrose cushion¹¹; fraction 4 was early endosome-enriched and corresponded to the 16%/10% sucrose interface; fraction 7 corresponded to the load of the gradient. High-speed pellet and supernatant fractions were prepared as in ref. 5. The affinity-purified antibodies raised against peptides derived from the C-terminal regions of rab5 and rab7 proteins were used at concentrations sufficient to detect only the overexpressed proteins (see also Fig. 2). Bands were visualized using ¹²⁵I-protein A and autoradiographs were scanned with the Ultrascan system (LKB).

affinity-purified antisera directed against either a rab7 C-terminal peptide (Fig. 3a) or an MSII polymerase-rab5 fusion protein⁵ (not shown). The presence of the rab5, rab7 and Nrab5₁₈₁Crab7₁₇₄ overexpressed proteins was also tested for in the other fractions of the flotation gradient (Fig. 3b). Rab5 cofractionated with early endosomes (fraction 4). Both rab7 and Nrab5₁₈₁Crab7₁₇₄ cofractionated with late endosomes (fraction 2). Little hybrid protein (8% of total) was detected in the early endosome fraction. As with overexpressed rab7, the specific association of Nrab5₁₈₁Crab7₁₇₄ protein with late endosomes was confirmed by immunoelectron microscopy analysis (data not shown). These data strongly argue against a nonspecific association of the overexpressed rab proteins to membranes other than early or late endosomes.

These results indicate that the C-terminal 34 residues of rab7 can target an N-terminal rab5 fragment of 181 residues to late endosomes. Neither the cysteine motif of rab7 nor adjacent sequences (up to 13 amino acids) can mediate this targeting by themselves. These results suggest that the targeting signal involves residues between position 174 and 195 of the rab7 protein.

We next examined whether a rab protein from the exocytic pathway could be targeted to an endocytic compartment by exchanging its C-terminal region. As a reporter molecule we used rab2, a protein localized to the intermediate compartment between the endoplasmic reticulum and the Golgi complex⁵. The overexpressed rab2 protein was barely detectable in both endosomal fractions (not shown). A hybrid protein consisting of 166 N-terminal rab2 residues fused with the 35 C-terminal amino acids of rab5 (Nrab2₁₆₆-Crab5₁₈₁) was detected only in the early endosome fraction (Fig. 3a), whereas the same reporter fragment fused to the C-terminal 35 amino acids of rab7 (Nrab2₁₆₆-Crab7₁₇₃) was found in the late endosome fraction. Although we cannot exclude the possibility that a minor fraction of these hybrid rab proteins is also associated with the exocytic pathway, these data indicate that the C-terminal regions of rab5 and rab7 can target a rab2 reporter fragment to early and late endosomes, respectively.

This study suggests that the specificity of localization for rab proteins is determined by the highly variable C-terminal sequence, as this region is necessary for correct targeting of the rab proteins tested here. Not all residues in this region are

involved in determining the specificity of localization. The C-terminal cysteine motif is required for membrane association^{5,14-17} but does not alone specify the targeting. Replacing the last 8 and 13 C-terminal residues of rab5 with the corresponding sequences from rab7 also did not affect localization of the hybrid protein to early endosomes. Although these residues could not influence targeting, their presence was, however, necessary for membrane association. A deletion of 10 amino acids in the C terminus of rab5 (rab5 $\Delta_{201-211}$) renders the protein cytosolic in spite of the presence of an intact cysteine motif (Fig. 3c), indicating that sequences adjacent to the cysteine motifs are also required. This suggests that the integrity of the C terminus is important for targeting and membrane association. Deletions within the C-terminus might perturb the correct folding of rab proteins or affect post-translational modification of the C-terminal cysteines.

Sequences next to the CAAX box are required for plasma membrane association of the ras proteins although it is not clear whether they also participate in targeting. Point mutations or deletion of a C-terminal polylysine sequence in p21^{K-ras} render this protein cytosolic²⁴ and membrane association of p21^{H-ras} (refs 19, 24) depends on palmitoylation of two cysteines immediately upstream of the CAAX motif.

We propose that rab proteins interact through their C-terminal hypervariable domain with specific receptors on the target membrane or with cytosolic proteins mediating the targeting event. Characterization of the machinery responsible for this process, which would include the enzymes catalysing the post-translational modifications on the C-terminal cysteines, will be crucial in order to understand how rab proteins function in controlling membrane traffic. Identification of the targeting signal on rab proteins should facilitate this task. □

Received 8 April; accepted 11 September, 1991.

1. Touchot, N., Chardin, P. & Tavittian, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8210-8214 (1987).
2. Buccì, C. et al. *Nucleic Acids Res.* **16**, 9979-9993 (1988).
3. Matsui, Y. et al. *J. Biol. Chem.* **263**, 11071-11074 (1988).
4. Zahraoui, A., Touchot, N., Chardin, P. & Tavittian, A. *J. Biol. Chem.* **264**, 12394-12401 (1989).
5. Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. *Cell* **62**, 317-329 (1990).
6. Chavrier, P., Vingron, M., Sander, C., Simons, K. & Zerial, M. *Molec. cell. Biol.* **10**, 6578-6585 (1990).
7. Gallwitz, D., Donath, C. & Sander, C. *Nature* **306**, 704-707 (1983).
8. Salminen, A. & Novick, P. *J. Cell* **49**, 527-538 (1987).
9. Segev, N., Mulholland, J. & Botstein, D. *Cell* **52**, 915-924 (1988).
10. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
11. Gove, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. *Cell* **64**, 915-925 (1991).
12. Goud, B., Zahraoui, A., Tavittian, A. & Saraste, J. *Nature* **345**, 553-556 (1990).
13. Fischer von Mollard, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1988-1992 (1990).
14. Willumsen, B. M., Norris, K., Papageorge, A. G., Hubbert, N. L. & Lowy, D. R. *EMBO J.* **3**, 2581-2585 (1984).

15. Deschenes, R. J. & Broach, J. R. *Molec. cell. Biol.* **7**, 2344-2351 (1987).
16. Molenaar, C. M. T., Prange, R. & Gallwitz, D. *EMBO J.* **7**, 971-976 (1988).
17. Walworth, N. C., Goud, B., Kastan Kabenell, A. & Novick, P. *J. EMBO J.* **8**, 1685-1693 (1989).
18. Gutierrez, L., Magee, A. I., Marshall, C. J. & Hancock, J. F. *EMBO J.* **8**, 1093-1098 (1989).
19. Hancock, J. F., Magee, A. I., Childs, J. E. & Marshall, C. J. *Cell* **57**, 1167-1177 (1989).
20. Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122-8126 (1986).
21. Zerial, M., Melancon, P., Schneider, C. & Garoff, H. *EMBO J.* **5**, 1543-1550 (1986).
22. Hopkins, C. R. *Cell* **35**, 321-330 (1983).
23. Schmid, S. L., Fuchs, R., Male, P. & Mellman, I. *Cell* **52**, 73-83 (1988).
24. Hancock, J. F., Paterson, H. & Marshall, C. J. *Cell* **63**, 133-139 (1989).
25. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. *Gene* **77**, 61-68 (1989).

ACKNOWLEDGEMENTS. We thank R. Jelinek, M. Stapleton and C. Walter for technical assistance, A. Sommerfeld for photographic work and R. Parton for immunoelectron microscopy analysis. We acknowledge N. Emans, S. Fuller, G. Griffiths and B. Hoflack for reading the manuscript. P.C. was recipient of an EMBL fellowship, J.-P. G. of a von Humboldt Stiftung fellowship.

Efficient production of functional mRNA mediated by RNA polymerase I in *Trypanosoma brucei*

Joost C. B. M. Zomerdijk, Rudo Kieft & Piet Borst

Division of Molecular Biology, the Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands

THE unicellular eukaryote *Trypanosoma brucei* evades the immune defence of its mammalian host by antigenic variation¹. The genes for variant-specific surface glycoproteins (VSGs) are expressed within large multicistronic transcription units². Mature messenger RNAs are produced by *trans*-splicing and polyadenylation³⁻⁵. A remarkable feature of the transcription of VSG genes is its insensitivity to the RNA polymerase II inhibitor α -amanitin⁶. This has led to the speculation that RNA polymerase I, normally only involved in the transcription of ribosomal RNA genes, also mediates expression of these surface antigen genes. In higher eukaryotes, however, transcripts produced by RNA polymerase I were found to be poor substrates for processing into mature mRNAs⁷⁻⁹. In contrast, we show here that the RNA polymerase I of *T. brucei* can mediate the efficient production of functional mRNA for neomycin phosphotransferase. This exceptional ability may be related to the unusual way in which pre-mRNAs are capped in trypanosomes. In most eukaryotes, mRNAs are modified at their 5' end by a capping activity associated with RNA polymerase II¹⁰; in trypanosomes, mRNAs acquire their 5'-cap from capped mini-exon donor RNA by *trans*-splicing³⁻⁵, a process that could be independent of the RNA polymerase producing the pre-mRNA.

The analysis of protein-coding transcripts produced by RNA polymerase I can be complicated by the production of transcripts initiating at RNA polymerase II, rather than polymerase I,

startpoints^{8,11}. We therefore aimed for a *neo*^r-gene integrated downstream of a non-transcribed spacer in the rDNA cluster, such that the endogenous rDNA promoter would drive the expression of a single *neo*^r-gene. To this end, we constructed the vector pRK20(+) depicted in Fig. 1a and introduced it in linear form into procyclic (insect form) *T. brucei* by electroporation^{12,13}. We obtained several independent stable transformants (at a frequency of about 1 in 10⁶ trypanosomes) with an integration at a single chromosomal locus coinciding with sequences homologous to the rDNA promoter.

Here we discuss the results of transformant T20, resistant to at least 2 mg G418 per ml. T20 expresses the *neo*^r-gene from a chromosome of ~2,000 kilobases (kb) (Fig. 1b). The physical map of the rRNA gene repeat unit is known¹⁴ (Fig. 1d, top) and this allowed us to predict the changes in the pattern of restriction fragments in genomic DNA after insertion of pRK20(+) (Fig. 1d, bottom), despite the multicopy nature of ribosomal promoter sequences in the *T. brucei* genome. Some of these changes are illustrated in Fig. 1c. For example, in DNA of T20 digested with *Eco*RI and *Bgl*II we detect with the ribosomal promoter probe the predicted fragments of 5.4 and 1.5 kb, the former also hybridizing to the *neo* probe. These fragments are not present in DNA from non-transformed trypanosomes (compare lanes 3 and 4), which only contains the multicopy 4.8-kb fragments derived from the rRNA gene repeats. All digests shown in Fig. 1c are compatible with a targeted insertion by homologous recombination of a single pRK20(+) plasmid into the rDNA promoter (Fig. 1d).

The sensitivity of *neo*^r-gene transcription to α -amanitin in T20 was determined by measuring transcription elongation in isolated nuclei. Transcription of the *neo*^r protein-coding gene is insensitive to α -amanitin (1 mg ml⁻¹), whereas the transcription of tubulin genes is >90% inhibited (Fig. 2a). Specific initiation of *neo* RNA synthesis by RNA polymerase I was analysed using RNase protection mapping of RNA 5' ends. The start site was directly compared with the endogenous rDNA start site. The antisense RNA probe, depicted in Fig. 2b, can

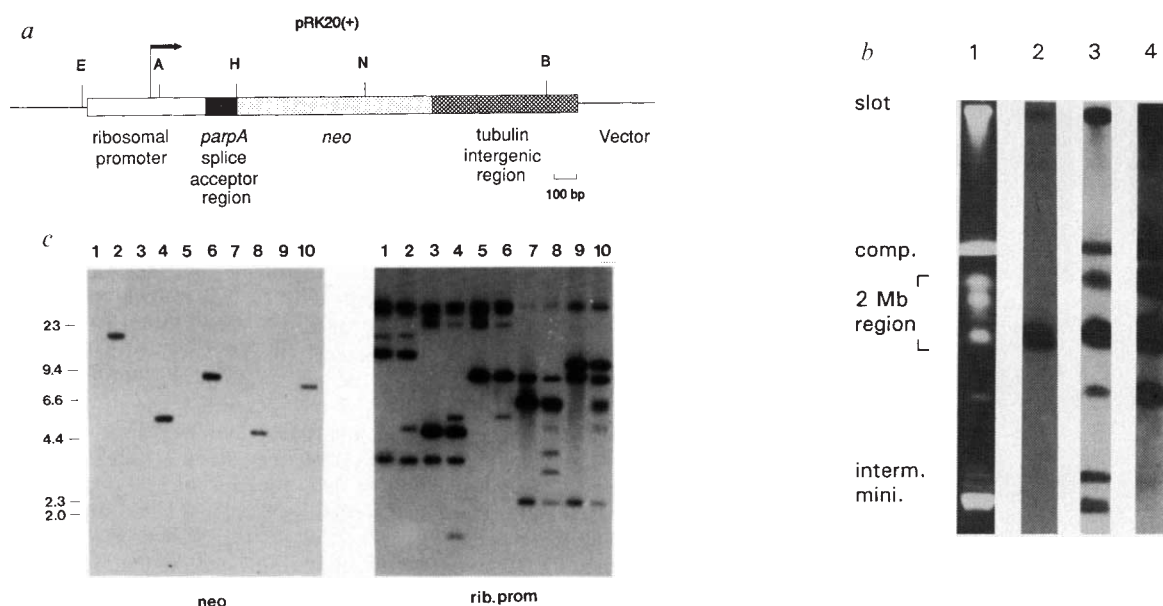
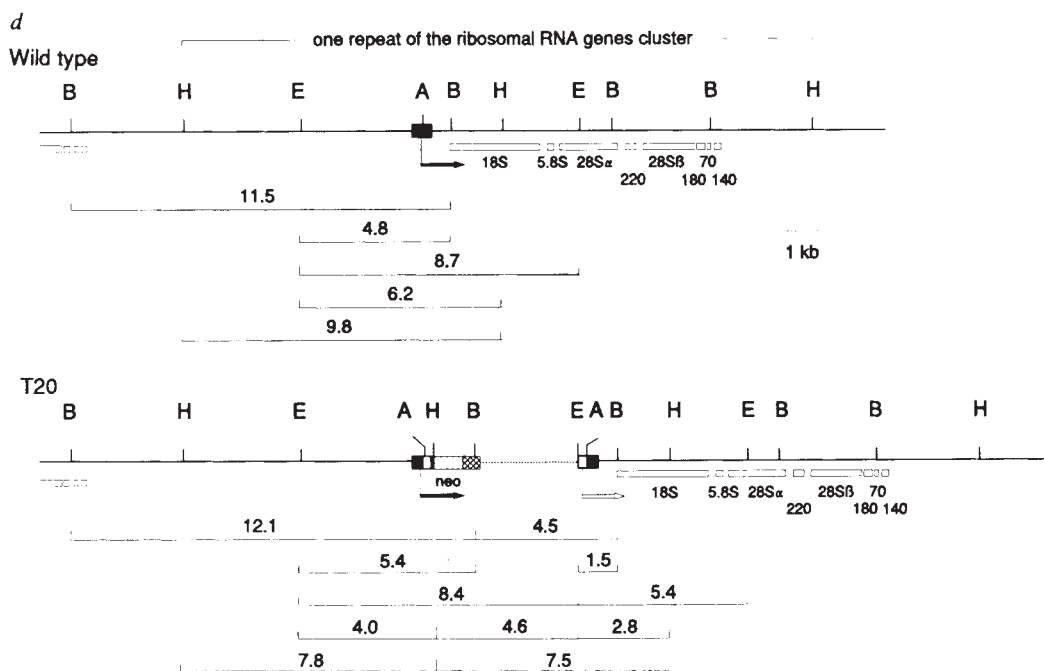


FIG. 1 Targeted insertion of the *neo*^r-gene into a rDNA cluster in *T. brucei*. **a**, Ribosomal promoter *neo*^r-gene chimaeric construct. The open box is the *T. brucei* rDNA promoter (519-bp *A**l**u**I* fragment) from -258 to +261 relative to the transcription start site^{14,24}. The hatched box represents a 130-bp fragment of the *T. brucei* procyclic or *parpA* 5' splice acceptor region to 113 bp upstream of the translation start codon of the *parpA* locus α -gene¹⁵. The stippled box is a 875-bp partial *Pst**I*-*Bam**H**I* fragment from plasmid pMC1POLA (Stratagene, USA) encompassing the bacterial *neo*^r-protein coding region. The cross-hatched box represents the intergenic region between the genes for β - and α -tubulin of *T. brucei*. The straight line is the pGEM3 vector (Promega). Restriction sites: A, *A**va**I*; B, *B**g**l**I*; E, *E**co**R**I*; H, *H**in**d**I**l**I*; N, *N**co**I*. **b**, The chromosome-sized DNA molecules of *T. brucei* T20 were separated in a pulsed-field gradient (PFG) gel²¹. The ethidium stained gel depicted in lane 1 was blotted and hybridized with a *neo*-probe (see stippled box in **a** and lane 2), a 519-bp *A**l**u**I* fragment encompassing the rDNA promoter probe (see open box in **a** and lane 3) and a 1.4-kb *B**g**l**I*-*H**in**d**I**l**I* 18S rRNA gene probe (see physical map in Fig. 1*d* and lane 4). The chromosomes in the size range of 2,000 kb, the intermediate chromosomes, and the minichromosomes are labelled as 2 megabase (Mb) region, interm. and mini. respectively. Very large chromosomes are not well separated and migrate in a compression zone, labelled comp. Not all loci that hybridize to the rDNA promoter encompass a complete rDNA repeat (compare lanes 3 and 4). The function and chromosomal context of these putative rDNA promoters outside of a rDNA cluster is not known. **c**, Southern blot analysis of *T. brucei* genomic DNA wild-type (lanes 1, 3, 5, 7, 9) and T20 (lanes 2, 4, 6, 8, 10), digested with the restriction enzymes *B**g**l**I* (lanes 1, 2), *E**co**R**I* and *B**g**l**I* (lanes 3, 4), *E**co**R**I* (lanes 5, 6), *E**co**R**I* and *H**in**d**I**l**I* (lanes 7, 8) and *H**in**d**I**l**I* (lanes 9, 10). The left blot was hybridized with the *neo*^r-gene probe, and the right with the rDNA promoter probe. *H**in**d**I**l**I* restriction fragments (sizes in kb) of phage λ DNA are indicated alongside the blots. The additional *H**in**d**I**l**I* fragments in the ribosomal promoter hybridization are due to DNA rearrangements unrelated to the integration event. **d**, Physical map of one repeat unit of the rRNA gene cluster in wild-type and in T20 *T.*



brucei genomic DNA containing a targeted insertion of pRK20(+) at the site of the rDNA promoter. The positions of the rRNA genes are indicated¹⁴. The black box with the arrow represents the endogenous rDNA promoter, whereas the open box with arrow represents the pRK20(+) plasmid-derived rDNA promoter. Symbols for the integrated pRK20(+) are as in **a**, except that pGEM vector DNA sequences are depicted as a dotted line. Hybridizing DNA fragments and their sizes in kilobases, corresponding to the digests and probes in **c** are shown. Restriction sites as in **a**. **METHODS.** 4 μ g of the 5-kb pRK20(+) plasmid DNA was linearized at the unique *A**va**I* site at +62 bp relative to the rDNA transcription start site, yielding DNA ends of 320 and 199 bp homologous to the rDNA promoter region, and introduced by electroporation into procyclic *T. brucei* 427¹². We obtained several cloned transformants by limiting dilution and selection on 20 μ g G418 per ml (IC_{50} for wild-type procyclics is 7 μ g ml⁻¹). DNA of wild-type and transformed *T. brucei* T20 cells, lysed *in situ* in 1% low melting point agarose, was separated in a pulsed-field gradient 0.5% agarose gel in 1 $\times$ TBE at 14 $^{\circ}$ C for 24 h at 10 V cm⁻¹, with the electric field switching every 180 s. Restriction fragments of genomic DNA were separated in 0.7% agarose 1 $\times$ TBE slab gels for 20 h at 1.5 V cm⁻¹. Gels were blotted and hybridized as described¹². Blots were washed in 0.1 $\times$ SSC at 65 $^{\circ}$ C.

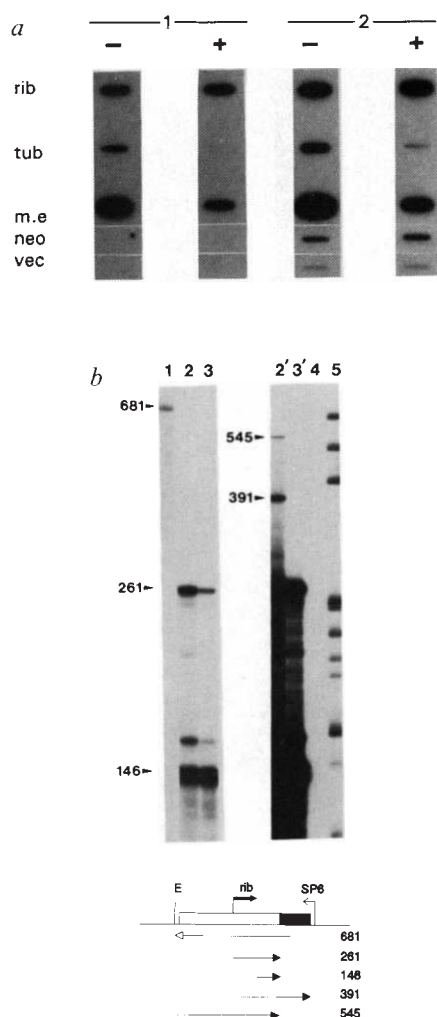


FIG. 2 Analysis of initiation and elongation of transcription in *T. brucei* T20 cells. **a**, Transcription of the *neo*^r-gene in isolated nuclei of wild-type (panel 1) and T20 (panel 2) *T. brucei* procyclics. Nascent RNA synthesized in isolated nuclei in the absence (–) or presence (+) of 1 mg α -amanitin per ml¹² was hybridized with slot-blot containing: rib, 4.8 kb *EcoRI*-*Bgl*III fragment overlapping the rDNA transcription start site (see Fig. 1d); tub, a 3.6 kb α - β tubulin repeat unit fragment; m.e. (mini-exon), pSPMEB5⁵; neo, 0.9 kb *neo*^r-gene from pRK20(+); vec, pGEM3 vector DNA (Promega). **b**, Use of RNase protection to map the start of transcription in the rDNA promoter at the site of a *neo*^r-gene insertion. Lane 1, an antisense uniformly labelled SP6 probe of 681 nucleotides, covering the rDNA promoter of 519 bp (open box) and *parpA* splice acceptor region of 130 bp (hatched box); lanes 2 and 2', protected fragments in 30 μ g total RNA from *T. brucei* T20; lanes 3 and 3', in 30 μ g total RNA from wild-type *T. brucei*; lane 4, in 30 μ g *Escherichia coli* tRNA. The proposed identities of the protected fragments (sizes in nucleotides) are schematically depicted underneath the gel. Lane 5, end-labelled pAT153 DNA digested with *Hpa*II. The left panel (lanes 2 and 3) is a 1-h exposure of the same gel exposed for 20 h and shown on the right (lanes 2' and 3'). The 5' end of a ribosomal RNA precursor at +115 (protected fragment of 146 nucleotides) is most likely an artefact of the nuclease protection analyses^{14,24}.

METHODS. Preparations of nuclei from trypanosomes and elongation of nascent RNA in nuclear run-ons are as described¹². [α -³²P]GTP was used as a tracer and radiolabelled nascent RNA was hybridized to slot blots, that were subsequently washed in 0.3 $\times$ SSC at 65 $^{\circ}$ C. The uniformly labelled antisense RNA probe, used in the RNase protection experiments, was synthesized *in vitro* with SP6 RNA polymerase and [α -³²P]CTP²², from the 674-kb *EcoRI*-*Hind*III fragment of pRK20(+) subcloned in pGEM3 (Promega), and hybridized with total trypanosome RNA, prepared in guanidinium isothiocyanate, for 16 h at 55 $^{\circ}$ C in 80% formamide, 40 mM PIPES, pH 6.4, 400 mM NaCl and 1 mM EDTA, followed by an RNase A (12 μ g) and T1 (3 U) digestion for 45 min at 30 $^{\circ}$ C. Protected fragments were visualized by electrophoresis in a 5% polyacrylamide, 8 M urea gel, followed by autoradiography.

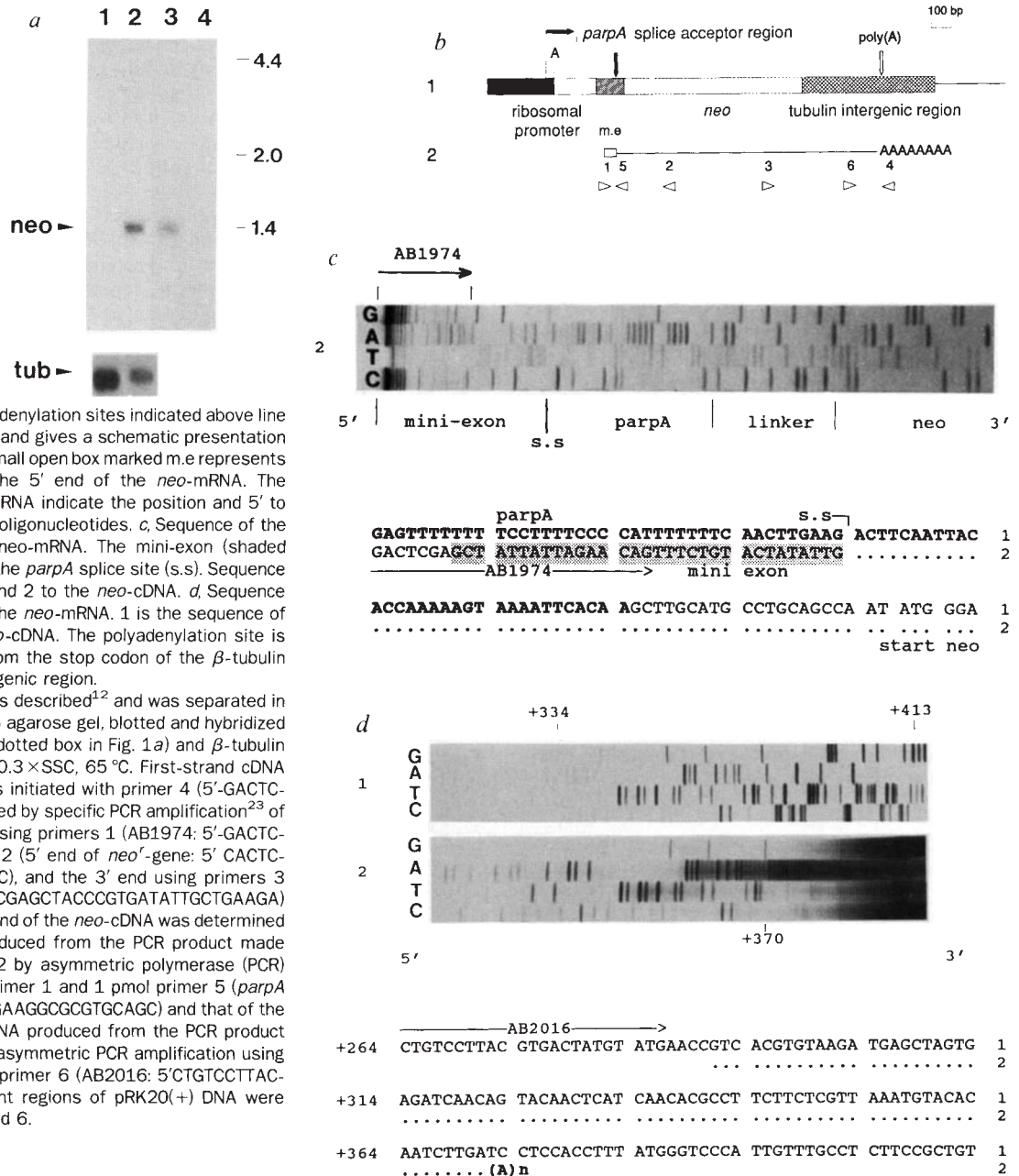
hybridize to endogenous pre-rRNA and to RNA derived from the integrated copy of pRK20(+). With this probe we observe a transcript of 391 nucleotides in RNA from T20 (Fig. 2b, lane 2'), not present in WT RNA (lane 3'). This transcript is 130 nucleotides longer than the abundant 261-nucleotide transcripts of the endogenous rDNA promoter. We deduce that this transcript is the chimaeric precursor RNA molecule synthesized at the integration site, initiating at the authentic transcription start site for RNA polymerase I of the rDNA promoter and overlapping the *parpA* splice acceptor region of 130 base pairs (bp). The other transcript, of 545 nucleotides, that is observed in T20 RNA is probably the result of runthrough transcription also seen in the run-on analyses (see Fig. 2a, vec.). This initiated from the rDNA promoter at the 5' junction of the genomic DNA to pRK20(+) DNA (see Fig. 1d, line 2, filled arrow), over the rDNA promoter that is reconstructed at the 3' junction as a consequence of the targeting event (open arrow). Taken together these results indicate that the *neo*^r-gene on the integrated copy of pRK20(+) is being transcribed from the rDNA promoter by RNA polymerase I.

In all eukaryotes analysed, including *T. brucei*, the processing of RNA synthesized by RNA polymerase II differs radically from that of pre-rRNA synthesized by polymerase I. We therefore examined the fate of *neo* transcripts produced by RNA polymerase I in our T20 cells. Figure 3a shows a 1.4-kb *neo*-hybridizing transcript present in steady-state RNA from T20 trypanosomes (lane 2) that is enriched in the polyadenylated RNA fraction (compare lanes 3 and 4). We made cDNAs of these transcripts and analysed the 5' and 3' ends following the strategy depicted in Fig. 3b. Figure 3c shows that the 5' end of the *neo*-mRNA carries the mini-exon (or spliced leader) sequence, that must have been added by *trans*-splicing to the authentic splice acceptor for the *parpA* gene¹⁵. Figure 3d shows that the poly(A) tail of the *neo* mRNA is at the site previously identified for polyadenylated β -tubulin mRNAs^{16,17}. These results show that the transcripts produced by RNA polymerase I are properly *trans*-spliced and polyadenylated. Processing must occur efficiently as the single *neo*^r-gene copy in the rDNA cluster yields more mRNA and a much higher resistance to G418 (>2 mg ml⁻¹) than a single *neo*^r-gene inserted in the tubulin gene cluster¹³ (about 70 μ g ml⁻¹; data not shown). This high efficiency has also been observed during transient expression of constructs containing the chloramphenicol acetyltransferase (CAT) gene. The CAT gene driven by the ribosomal promoter yielded as much CAT as the CAT gene driven by the VSG or procyclin gene promoters²⁴. Hence, our results prove that RNA polymerase I can mediate the efficient expression of protein-coding genes in trypanosomes.

We are aware of only two examples of naturally occurring protein-coding genes in nuclear rDNA of eukaryotes. Xiong and Eickbush¹⁸ described a retrotransposon, coding for a putative reverse transcriptase and endonuclease, in 28S rDNA of the silkworm *Bombyx mori* and Muscarella and Vogt¹⁹ identified a mobile group I intron encoding an endonuclease in the nuclear rDNA of *Physarum polycephalum*. In both cases the encoded protein is probably required only in low amounts and the structure of the putative mRNAs and their genesis in the splicing reactions are not yet known. Trypanosomes therefore seem unusual among eukaryotes in being able to use polymerase I transcripts for efficient mRNA synthesis. Why other eukaryotes lack this ability is not known. Possible explanations include the segregation of nascent polymerase I transcripts in the nucleolus, making them inaccessible to pre-mRNA processing enzymes, or a tight linkage of essential steps in pre-mRNA processing to RNA polymerase II transcription. It seems unlikely that segregation is at fault, as rRNA synthesis in trypanosomes is also concentrated in one spot in the nucleus (unpublished observations). There is, however, a major difference in pre-mRNA processing. In most eukaryotes, each pre-mRNA is capped by

FIG. 3 Analysis of *neo^r*-gene transcripts. *a*, *neo^r*-gene transcripts in 5 µg *T. brucei* total RNA from wild-type (lane 1) and T20 (lane 2) cells, and in 0.5 µg poly(A)⁺ (lane 3) and 5 µg poly(A)⁻ fractionated RNA (lane 4) from T20 *T. brucei*. *b*, Schematic presentation of the integrated pRK20(+) DNA construct (as depicted in Fig. 1*d*) with the black box representing the genomically derived rDNA promoter sequences, the open box vector-derived rDNA sequences and the mini-exon splice acceptor and polyadenylation sites indicated above line 1. Line 2 is aligned with line 1 and gives a schematic presentation of the *neo^r*-gene mRNA. The small open box marked m.e. represents the mini-exon sequence at the 5' end of the *neo*-mRNA. The arrowheads underneath the mRNA indicate the position and 5' to 3' direction of synthetic deoxyoligonucleotides. *c*, Sequence of the 5' end of the cDNA of the *neo*-mRNA. The mini-exon (shaded sequence) is *trans*-spliced to the *parpA* splice site (s.s). Sequence 1 corresponds to pRK20(+) and 2 to the *neo*-cDNA. The polyadenylation site is indicated at position +370 from the stop codon of the β -tubulin gene in the β - α tubulin intergenic region. *d*, Sequence of the 3' end of the cDNA of the *neo*-mRNA. 1 is the sequence of pRK20(+) and 2 is of the *neo*-cDNA. The polyadenylation site is indicated at position +370 from the stop codon of the β -tubulin gene in the β - α tubulin intergenic region.

METHODS. RNA was isolated as described¹² and was separated in a denaturing formaldehyde 1% agarose gel, blotted and hybridized to the *neo^r*-gene probe (*neo*; dotted box in Fig. 1*a*) and β -tubulin probe. Blots were washed in 0.3×SSC, 65 °C. First-strand cDNA synthesis²¹ of *neo*-mRNA was initiated with primer 4 (5'-GACTC-GAGTCGACATCGA(T)₁₇), followed by specific PCR amplification²³ of the 5' end of the *neo*-mRNA using primers 1 (AB1974: 5'-GACTC-GAGCTATTATTAGAACAG) and 2 (5' end of *neo*-cDNA: 5' CACTC-GAGTCGTGGCCAGCCACGATAGC), and the 3' end using primers 3 (3' end of *neo*-cDNA: 5'-GACTCGAGCTACCGTGATATTGCTGAAGA) and 4. The sequence of the 5' end of the *neo*-cDNA was determined with single-stranded DNA produced from the PCR product made with primer 1 and 2 by asymmetric polymerase (PCR) amplification using 50 pmol primer 1 and 1 pmol primer 5 (*parpA* splice acceptor region: 5'-CTCGAAGGCGCGTGCAGC) and that of the 3' end with single-stranded DNA produced from the PCR product made with primer 3 and 4 by asymmetric PCR amplification using 50 pmol primer 4 and 1 pmol primer 6 (AB2016: 5'-CTGTCCCTTAC-GTGACTATGTATG). The relevant regions of pRK20(+) DNA were sequenced using primers 2 and 6.



guanylyl transferase immediately after initiation of transcription²⁰. Capping seems to be closely associated with RNA polymerase II¹⁰, and this may prevent capping of transcripts generated by polymerase I. In contrast, all mRNAs in *T. brucei* acquire their 5'-cap by *trans*-splicing. Cap acquisition is ensured by

incorporating a *trans*-splice acceptor site in the precursor transcript. This may make capping independent of the RNA polymerase producing the pre-mRNA and explain the talent of *T. brucei* to process RNA polymerase I generated transcripts into functional mRNAs. □

Received 22 July; accepted 4 September 1991.

1. Cross, G. A. M. *A. Rev. Immun.* **8**, 83-110 (1990).
2. Johnson, P. J., Kooter, J. M. & Borst, P. *Cell* **51**, 273-281 (1987).
3. Murphy, W. J., Watkins, K. P. & Agabian, N. *Cell* **47**, 517-525 (1986).
4. Sutton, R. E. & Boothroyd, J. C. *Cell* **47**, 527-535 (1986).
5. Laird, P. W., Zomerijk, J. C. B. M., De Korte, D. & Borst, P. *EMBO J.* **6**, 1055-1062 (1987).
6. Kooter, J. M. & Borst, P. *Nucleic Acids Res.* **12**, 9457-9472 (1984).
7. Grummt, I. & Skinner, J. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 722-726 (1985).
8. Smale, S. T. & Tjian, R. *Molec. cell. Biol.* **5**, 352-362 (1985).
9. Grimaldi, G. & Di Nocera, P. *Nucleic Acids Res.* **14**, 6417-6432 (1986).
10. Jove, R. & Manley, J. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5842-5846 (1982).
11. Lopata, M. A., Cleveland, D. W. & Sollner-Webb, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6677-6681 (1986).
12. Zomerijk, J. C. B. M. *et al.* *EMBO J.* **9**, 2791-2801 (1990).
13. Ten Asbroek, A. L. M. A., Ouellette, M. & Borst, P. *Nature* **348**, 174-175 (1990).
14. White, T. C., Rudenko, G. & Borst, P. *Nucleic Acids Res.* **14**, 9471-9489 (1986).

15. Clayton, C. E. *et al.* *Molec. cell. Biol.* **10**, 3036-3047 (1990).
16. Sather, S. & Agabian, N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5695-5699 (1985).
17. Imboden, M., Blum, B., De Lange, T., Braun, R. & Seebeck, T. *J. molec. Biol.* **166**, 393-402 (1986).
18. Xiong, Y. & Eickbush, T. H. *Cell* **55**, 235-246 (1988).
19. Muscarella, D. E. & Vogt, V. M. *Cell* **56**, 443-454 (1989).
20. Salditt-Georgieff, M., Harpold, M., Cheng-Kiang, S. & Darnell, J. E., Jr. *Cell* **19**, 69-78 (1980).
21. Zomerijk, J. C. B. M., Kieft, R., Duyndam, M., Shiels, P. G. & Borst, P. *Nucleic Acids Res.* **19**, 1359-1368 (1991).
22. Zinn, K., DiMaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).
23. Saiki, R. K. *et al.* *Science* **239**, 487-491 (1988).
24. Zomerijk, J. C. B. M., Kieft, R., Shiels, P. G. & Borst, P. *Nucleic Acids Res.* **19**, 5153-5158 (1991).

ACKNOWLEDGEMENTS. We thank A. ten Asbroek, L. Eijdens, J. Gommers-Ampt, C. Lincke, J. Smit, F. Baas, R. Plasterk, J. Russell and A. Schinkel for interest and for reading the manuscript. This work was supported in part by a grant from the Netherlands Foundation for Chemical Research (SON), with financial aid from the Netherlands Organization for Scientific Research (NWO).

The *vsr* gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease

Frank Hennecke, Harald Kolmar, Kerstin Bründl & Hans-Joachim Fritz

Institut für Molekulare Genetik, Georg-August-Universität Göttingen, Grisebachstrasse 8, D-3400 Göttingen, Germany

IN *Escherichia coli* K-12, the Dcm methyltransferase catalyses methylation of the inner cytosine residue in the sequence $CC^A/TGG^{1,2}$. Hydrolytic deamination of 5-methylcytosine bases in DNA leads to thymine residues, and hence to T/G mismatches, pre-mutagenic DNA lesions consisting of two natural DNA constituents and thus devoid of an obvious marker of the damaged DNA strand. These mismatches are corrected by the VSP repair pathway, which is characterized by very short patches of DNA repair synthesis^{3,4}. It depends on genes *vsr*⁵ and *polA*⁶ and is strongly stimulated by *mutL* and *mutS*⁷⁻⁹. The *vsr* gene product (Vsr; M_r 18,000) was purified and characterized as a DNA mismatch endonuclease, a unique and hitherto unknown type of enzyme. Vsr endonuclease nicks double-stranded DNA within the sequence CT^A/TGN or NT^A/TGG next to the underlined thymine residue, which is mismatched to 2'-deoxyguanosine. The incision is mismatch-dependent and strand-specific. These results illustrate how Vsr endonuclease initiates VSP mismatch repair.

For overproduction of Vsr protein in *E. coli*, an in-frame fusion gene of the *bla* (β -lactamase) and *vsr* genes has been

constructed (Fig. 1a). The crude mixture of total cellular protein obtained on expression of the chimaeric gene was purified, exploiting binding of β -lactamase to immobilized phenylboronate¹⁶ (Fig. 1b). Oligonucleotide duplex FH5/FH2, which contains a T/G mismatch in the sequence context of the phage λ *cI am6* mutation^{3,4}, was incubated with purified β -lactamase-Vsr fusion protein (Fig. 2). Electrophoretic mobilities of reaction products relative to marker oligonucleotides suggest endonucleolytic cleavage at the site indicated by the arrow. The association of the catalytic activity with the Vsr moiety of the β -lactamase-Vsr fusion protein was confirmed by the observation that a β -lactamase-*gal* repressor fusion protein, purified by the same procedure (Fig. 1b, lane 5), failed to mediate the cleavage reaction. Results obtained with either type of fusion protein were independent of the choice of host strain BMH71-18/*mutS* or MG15(Δ *vsr*). This excludes any significant contribution of potentially copurified trace amounts of MutS or, respectively, Vsr protein to the activity observed. Enzymatic activities of homogeneous fusion protein and the product mixture of IgA protease cleavage¹⁴ (Fig. 1b, lane 4) were indistinguishable (data not shown). This congruence of functional characteristics, together with the favourable physical properties of the β -lactamase-Vsr fusion protein prompted us to use the latter in the qualitative assessment of mode of action.

Several specialized DNA mismatch repair mechanisms involve strand cleavage in a sequential fashion: hydrolysis of the glycosidic bond of one of the mismatched nucleotides^{19,20}, followed by AP-endonuclease-mediated incision. By contrast, Vsr-mediated reaction seems to proceed by direct endonucleolytic attack (Fig. 2); nevertheless, we investigated the possibility of release of a thymine base from duplex FH5/FH2, as well as cleavage of a duplex containing an apyrimidinic site in

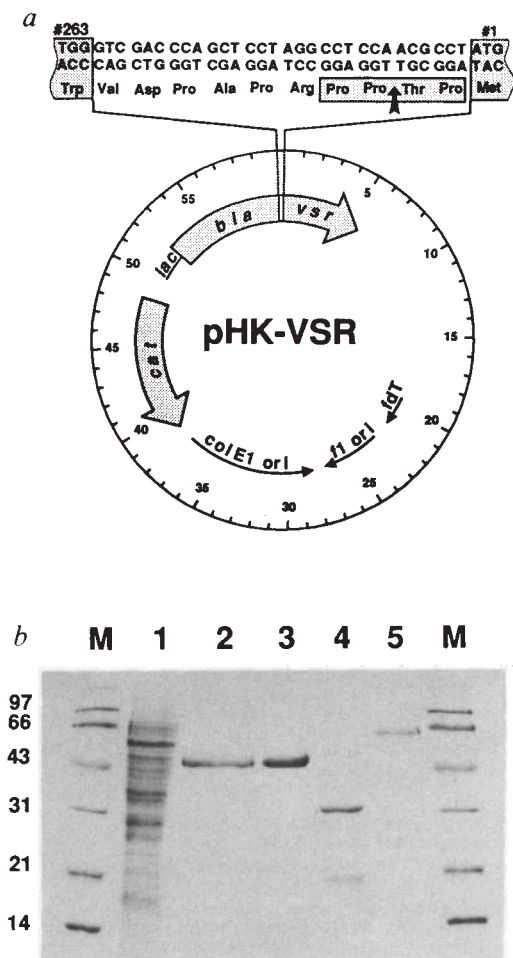
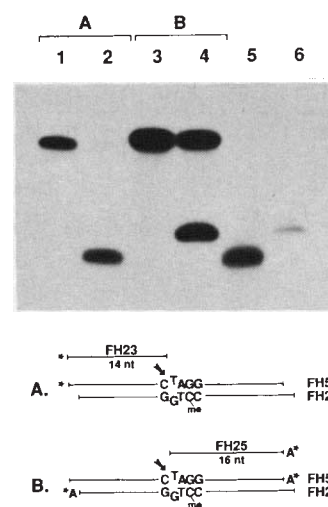


FIG. 1 Production of β -lactamase-Vsr fusion protein. **a**, Expression vector pHK-VSR, which has been constructed from phasmid pHKREI (H.K., K.B. and H.-J.F., unpublished) by several rounds of cloning and oligonucleotide directed mutagenesis using standard techniques^{10,11} contains the *vsr* gene^{5,12} (ranging from nucleotide pairs 1,697 to 2,167 according to GenEMBL accession no. X13330) fused to the *E. coli* β -lactamase coding sequence¹³ (ranging from nucleotide pairs 2,546 to 1,689 according to GenEMBL accession no. VB0040 with nucleotides 2,525 to 2,478, corresponding to codons -16 to -1 of the β -lactamase signal sequence, deleted). Both genes are connected by a synthetic 30-base-pair adaptor encompassing the recognition sequence of IgA protease from *Neisseria gonorrhoeae*¹⁴ (Pro-Pro-Thr-Pro). The IgA protease cleavage site is indicated by an arrow. In pHK-GalR, the *gal* repressor coding sequence¹⁵ (ranging from nucleotide pairs 2,378 to 1,119 according to GenEMBL accession no. Y01614) is fused in frame to the *bla* gene. Both fusion genes are expressed under *lac* promoter control. **b**, Analysis of various stages in the purification of β -lactamase-Vsr fusion protein by SDS-polyacrylamide gel electrophoresis (15% gel; Coomassie staining). Lane 1, crude lysate of induced *E. coli* BMH 71-18/*mutS*¹⁷ harbouring plasmid pHK-VSR; lane 2, after passage through phenylboronate affinity column¹⁶; lane 3, fusion protein after anion-exchange chromatography; lane 4, IgA protease cleaved; lane 5, affinity purified β -lactamase-*gal* repressor fusion protein; M: molecular weight markers with sizes indicated ($M_r \times 10^{-3}$). METHODS. *E. coli* BMH 71-18/*mutS* harbouring pHK-VSR was grown overnight in 1 litre rich medium at 25 °C with chloramphenicol (25 mg l⁻¹) and IPTG (160 μ M) added. After centrifugation at 3,200g, cells were resuspended in 20 ml 20 mM triethanolamine-HCl, 0.5 M NaCl, pH 7.0, lysed by passage through a French pressure cell at 16,000 p.s.i. and centrifuged at 9,000g. From the supernatant, the fusion protein was purified essentially as described¹⁶ using a phenylboronate sepharose affinity column with hydrophobic spacer arms. Typically, 1 mg of fusion protein was obtained. β -lactamase-Vsr containing fractions were diluted 10-fold in 75 mM Tris-HCl pH 8.8 and subjected to anion-exchange chromatography (Pharmacia mono Q HR 5/5 column) in 75 mM Tris-HCl, pH 8.8, with a 0 to 1 M NaCl gradient. The fusion protein eluted at 240 mM NaCl. IgA protease cleavage was performed by overnight incubation at 15 °C (200-fold molar excess of fusion protein). Identical growth conditions and purification procedures were used for isolation of the β -lactamase-*gal* repressor fusion protein. Both the β -lactamase-Vsr and the β -lactamase-*gal* repressor fusion protein were also produced and purified to the same level of homogeneity applying the identical procedure to accordingly transformed strain MG15 (Δ *vsr*).

FIG. 2 Oligonucleotide nicking assay. Oligonucleotides FH5, d(TACTTGGCT-TATCCTAGGAATCTGTGCGAG), and FH2, d(TCTGCGACAGATTC³²CTGGGATAAGCCAAGT), were chemically synthesized from phosphoramidite precursors¹⁸ and used for oligonucleotide duplex formation (³²P: DNA 5-methylcytosine residue). Sequences are derived from nucleotides 87 to 115 of the phage λ *cl* gene. They comprise the *dcm* methylation sequence which is also the site of the *cl* *am6* mutation^{3,4}. FH5 corresponds to the coding strand of *cl* *am6*, FH2 to the non-coding strand of wild type *cl*. Bold-faced lettering indicates the position of the original *dcm* methylation sequence. Lane 1, oligonucleotide duplex FH5/FH2 with oligonucleotide FH5 labelled with ³²P at its 5'-terminus; lane 2, same oligonucleotide duplex after incubation with β -lactamase-Vsr fusion protein; lane 3, oligonucleotide duplex FH5/FH2 labelled at both 3'-termini; lane 4, same oligonucleotide duplex after incubation with β -lactamase-Vsr fusion protein; lane 5, marker oligonucleotide FH23, d(TACTTGGCTTATCC), corresponding to the 5'-terminal 14 nucleotides of oligonucleotide FH5, radioactively phosphorylated at its 5'-terminus; lane 6, marker oligonucleotide FH25, d(TAGGAATCTGTGCGAG), corresponding to the 3'-terminal 16 nucleotides of oligonucleotide FH5, elongated at its 3'-terminus by one ³²P-labelled dAMP residue. The radioactively labelled double-stranded oligonucleotides as well as the marker oligonucleotides used are shown schematically below the autoradiograph. Positions of radioactive label are indicated by an asterisk. The arrow points to the site of cleavage by Vsr. Cleavage assays shown here were performed with β -lactamase-Vsr fusion protein isolated from BMH 71-18/*mutS* as the production host; identical results were obtained with the same protein isolated from MG15 (compare legend to Fig. 1).

METHODS. Oligonucleotides FH5 and FH2 were labelled with ³²P to a specific activity of approximately 2 to 4 nCi fmol⁻¹ (74 to 148 Bq fmol⁻¹). 5'-labelling was performed by incubation of oligonucleotide FH5 with [γ -³²P]ATP in the presence of T4 polynucleotide kinase. After removal of excess [γ -³²P]ATP by gel filtration through sephadex G-25, the oligonucleotide was annealed with a 20-fold excess of the non-radioactive complementary oligonucleotide FH2. For 3'-labelling, equimolar amounts of oligonucleotides FH5 and FH2 were annealed by heating to 80 °C and cooling to room temperature. The resulting oligonucleotide duplex was labelled at both 3'-ends by incubation



with Klenow fragment of DNA polymerase I and [α -³²P]dATP. Excess [α -³²P]dATP was removed by gel filtration. The respective radiolabelled double-stranded oligonucleotide (10 fmol) was incubated with 100 ng of β -lactamase-Vsr fusion protein in 25 mM HEPES-KOH, pH 7.9, 0.01 mM ZnCl₂, 0.5 mM dithiothreitol at 30 °C for 15 min in a total volume of 20 μ l. After phenol/chloroform extraction, the reaction mixture was separated by denaturing gel electrophoresis (15% polyacrylamide, 8 M urea) followed by autoradiography. The metal cofactor is Mg²⁺ tightly bound to the enzyme and thus carried through the entire purification procedure, rather than Zn²⁺ added with the incubation buffer (Wolfgang Gläsner, unpublished). Substitution of ZnCl₂ by MgCl₂ in the incubation buffer did not result in a detectable difference in activity.

place of the mismatched T residue, invariably with negative results. Direct evidence against a reaction proceeding by glycosylase attack on the mismatched T was obtained by nucleotide-sequence analysis of the 3'-terminal cleavage product using the chemical degradation method²¹. This analysis confirmed the presence of a thymine base at the 5'-end (data not shown).

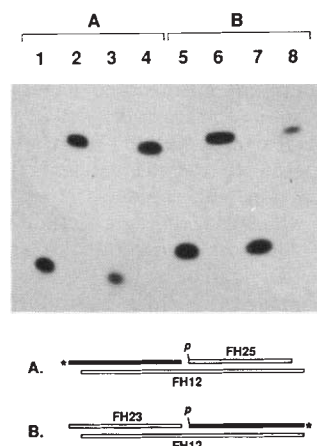
To determine the exact chemical nature of the termini generated by Vsr, radioactively labelled 5'- and 3'-terminal cleavage products were isolated and used as substrates for ligation reactions with defined oligonucleotides (Fig. 3). Efficient ligation of

the 5'-terminal cleavage product was achieved with 5'-phosphorylated oligonucleotide FH25, which contains the thymidine residue in question at its 5' terminus (experiment A). On the other hand, the 3'-cleavage product could smoothly be ligated to oligonucleotide FH23, indicating the presence of a 5'-terminal phosphate group (experiment B). Clearly, Vsr-mediated cleavage occurs in an endonucleolytic fashion immediately 5' to the mismatched thymine residue producing a 3'-hydroxyl and a 5'-phosphate group.

Substrate requirements were further investigated by incuba-

FIG. 3 Ligation assay. Isolated 5'-terminal Vsr cleavage product of FH5 and 5'-phosphorylated oligonucleotide FH25 (see legend to Fig. 2) were annealed to the orienting template oligonucleotide FH12, d(TCTGCGACAGATTC³²CTAGGATAAGCCAAGT). Note that oligonucleotide FH12, unlike FH2, corresponds to the non-coding strand of *cl* *am6*. Hence, no mismatch is formed in the resulting oligonucleotide duplex (situation A). Likewise, isolated 3'-terminal Vsr cleavage product of FH5 and oligonucleotide FH23 were annealed to FH12 to create situation B. Solid boxes represent the 5'- or 3'-terminal cleavage product, respectively; the radioactive label is indicated by an asterisk. Open boxes represent the synthetic oligonucleotides employed, P represents a 5'-phosphate group. After incubation with T4 DNA ligase and ATP, the mixtures were subjected to denaturing gel electrophoresis (15% polyacrylamide, 8 M urea). Situation A: lane 1, 5'-terminal cleavage product; lane 2, oligonucleotide duplex A after incubation with T4 DNA ligase; lane 3, marker oligonucleotide FH23, 5'-labelled with ³²P; lane 4, marker oligonucleotide FH5, 5'-labelled (see Fig. 2). Situation B: lane 5, 3'-terminal cleavage product; lane 6, oligonucleotide duplex B after incubation with T4 DNA ligase; lane 7, marker oligonucleotide FH25, non-radioactively phosphorylated at the 5'-terminus and 3'-labelled (see Fig. 2); lane 8, marker oligonucleotides FH5/FH2 labelled at both 3'-termini.

METHODS. Oligonucleotide duplex FH5/FH2 labelled radioactively at the 5'- or 3'-ends, respectively, was incubated with β -lactamase-Vsr fusion protein. After phenol/chloroform extraction, samples were subjected to denaturing gel electrophoresis followed by autoradiography (see Fig. 2). The bands representing radioactive 5'- or 3'-terminal cleavage products of oligonucleotide FH5, respectively, were excised and the DNA was eluted from the gel



slices by overnight incubation with water at 65 °C. Gel pieces were removed by centrifugation, DNA was subsequently recovered by ethanol precipitation. Around 20 fmol of either cleavage product were hybridized to 1 pmol of oligonucleotide FH12 together with the respective ligation partner (100 pmol).

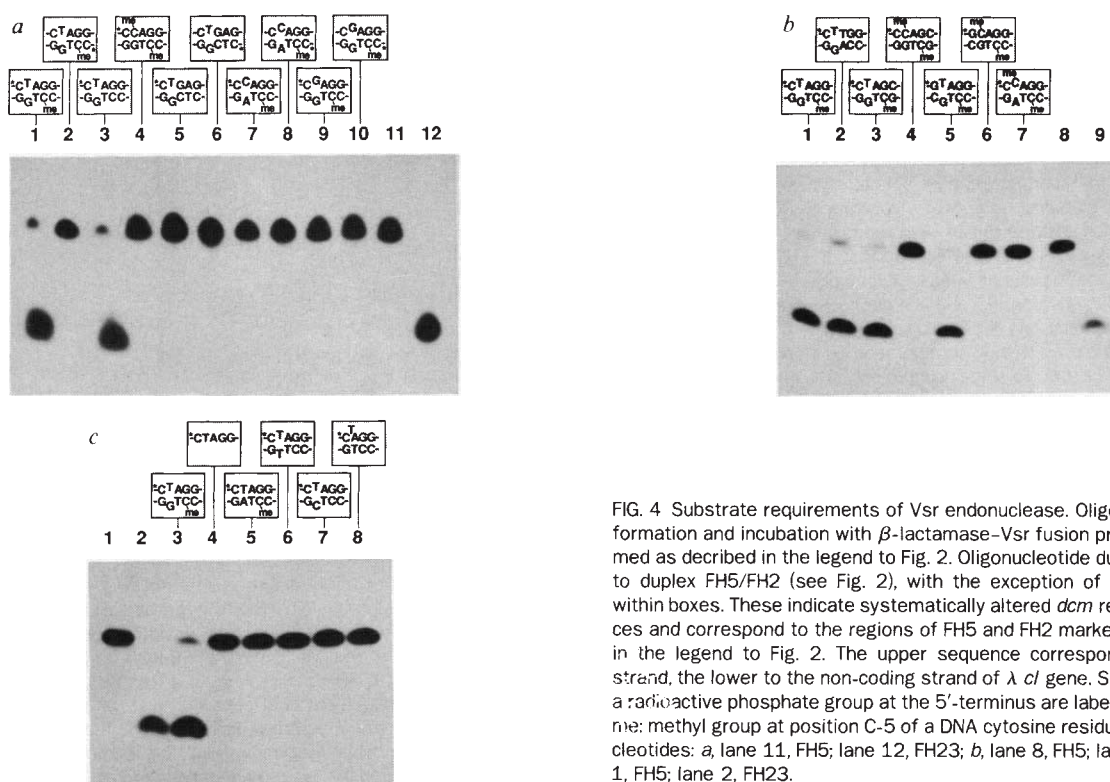


FIG. 4 Substrate requirements of Vsr endonuclease. Oligonucleotide duplex formation and incubation with β -lactamase-Vsr fusion protein were performed as described in the legend to Fig. 2. Oligonucleotide duplexes correspond to duplex FH5/FH2 (see Fig. 2), with the exception of sequences shown within boxes. These indicate systematically altered *dcm* recognition sequences and correspond to the regions of FH5 and FH2 marked by bold lettering in the legend to Fig. 2. The upper sequence corresponds to the coding strand, the lower to the non-coding strand of λ *cl* gene. Strands possessing a radioactive phosphate group at the 5'-terminus are labelled by an asterisk; me: methyl group at position C-5 of a DNA cytosine residue. Marker oligonucleotides: a, lane 11, FH5; lane 12, FH23; b, lane 8, FH5; lane 9, FH23; c, lane 1, FH5; lane 2, FH23.

tion of oligonucleotide duplexes comprising systematically altered sequences near the scissile phosphodiester bond with β -lactamase-Vsr fusion protein (Fig. 4). VSP repair was discovered as a genetic phenomenon associated with *dcm* methylation sites³. Accordingly, changing the relevant sequence from CTAGG to CTGAG prevents incision (Fig. 4a, lanes, 5, 6), whereas substitution of the central A-T base pair by T-A has no effect (Fig. 4b, lanes 1, 2). In agreement with genetic data²², one out of two peripheral nucleotide pairs of the target sequence can be changed without abolishing endonucleolytic cleavage by Vsr (Fig. 4b, lanes 3–6). The presence of a 5-methyl group on the inner cytosine residue of the uncleaved strand of the substrate sequence is not required (Fig. 4a, lane 3). Homoduplex DNA was resistant to Vsr action (Fig. 4a, lane 4; Fig. 4b, lanes 4, 6; Fig. 4c, lane 5), as were heteroduplex DNAs containing various different base–base mismatches other than T–G (Fig. 4a, lanes 7–10; Fig. 4b, lane 7, Fig. 4c). The endonucleolytic attack on the T–G mismatch by Vsr is specifically directed against the DNA strand containing the mismatched T (Fig. 4a, lanes 1, 2).

In summary, Vsr endonuclease specifically recognizes a T–G mismatch in a special context related to the *dcm* methylation sequence and incises 5' to the mismatched T residue. Identification of this reaction with the decisive initial step of VSP repair is in accord with all available genetic evidence^{7–9}, even in subtle details of substrate recognition²². Genetic work with *polA* mutants⁶ suggests that once the Vsr cut is made, DNA polymerase I will remove the fraying thymidine residue with its 5'→3' exonuclease activity and commence repair synthesis in the usual template-directed fashion, with the short synthesis tracts typical for that enzyme²³. DNA ligase can safely be assumed to seal the DNA strand on which repair has been performed. Elucidation of the mechanism by which gene products MutL and MutS stimulate the process^{7–9} awaits further experiments.

In *E. coli* K-12, recognition sequences of *Dcm* methyltransferase are hotspots of spontaneous mutagenesis²⁴ despite the existence of VSP repair, hence, one has to expect an increased occurrence of sequence polymorphisms at such sites. In the course of heteroduplex formation during genetic recombination,

a correspondingly increased occurrence of mismatches that are Vsr substrates must be expected. Such mismatches could be recognized and processed by Vsr endonuclease with the initiation of a local repair event. Additional mismatches located in the vicinity could be passively corepaired in the same short pulse of DNA repair synthesis promoted by DNA polymerase I. The result would be unidirectional transfer of information along a very short patch of the genome. It is thus conceivable that Vsr activity provides a paradigm beyond *E. coli* and the prokaryotes for the role of DNA mismatch endonucleases in patchwise gene conversion as implied, for instance, in the diversification of chicken immunoglobulin genes^{25,26}. □

Received 8 July; accepted 5 September 1991.

- Märinus, M. G. *Molec. gen. Genet.* **127**, 47–55 (1973).
- Schlagmann, S., Hattmann, S., May, M. S. & Berger, L. *J. Bact.* **126**, 990–996 (1976).
- Lieb, M. *Molec. gen. Genet.* **191**, 118–125 (1983).
- Lieb, M. *Molec. gen. Genet.* **199**, 465–470 (1985).
- Shoail, A., Lieb, M., Dar, M. & Bhagwat, A. S. *J. Bact.* **172**, 4214–4221 (1990).
- Dzidic, S. & Radman, M. *Molec. gen. Genet.* **217**, 254–256 (1989).
- Zell, R. & Fritz, H.-J. *EMBO J.* **6**, 1809–1815 (1987).
- Lieb, M. *J. Bact.* **169**, 5241–5246 (1987).
- Jones, M., Wagner, R. & Radman, M. *J. molec. Biol.* **194**, 155–159 (1987).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367–382 (1987).
- Hanck, T., Gerwin, N. & Fritz, H.-J. *Nucleic Acids Res.* **17**, 5844 (1989).
- Stanssens, P., Opsomer, C., McKeown, Y. M., Kramer, W., Zabeau, M. & Fritz, H.-J. *Nucleic Acids Res.* **17**, 4441–4454 (1989).
- Pohlner, J., Halter, R., Beyreuther, K. & Meyer, T. F. *Nature* **325**, 458–462 (1987).
- W. Wilcken-Bergmann, B. & Müller-Hill, B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2427–2431 (1982).
- Cartwright, S. J. & Waley, S. G. *Biochem. J.* **211**, 505–512 (1984).
- Kramer, B., Kramer, W. & Fritz, H.-J. *Cell* **38**, 879–887 (1984).
- Beaucage, S. L. & Caruthers, M. H. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
- Au, K. G., Clark, S., Miller, J. H. & Modrich, P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8877–8881 (1989).
- Wiebauer, K. & Jiricny, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5842–5845 (1990).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
- Lieb, M., Allen, E. & Read, D. *Genetics* **114**, 1041–1060 (1986).
- Bambara, R. A., Uemura, D. & Choi, T. *J. biol. Chem.* **253**, 413–423 (1978).
- Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. *Nature* **274**, 775–780 (1978).
- Reynaud, C.-A., Anquez, V., Grimal, H. & Weill, J.-C. *Cell* **48**, 379–388 (1987).
- Wysocki, L. J. & Gefter, M. L. *A. Rev. Biochem.* **58**, 509–531 (1989).

ACKNOWLEDGEMENTS. We wish to thank W. Kramer for suggestions. This work was supported by Deutsche Forschungsgemeinschaft through Forschungsschwerpunkt Gentechnologie. H. Kolmar is recipient of a 'Kekulé' fellowship from Fonds der Chemischen Industrie.